ASSESSMENT AND RANKINGS EFFORTS: THE EFFECT ON INSTITUTIONAL AND PROGRAM-LEVEL CHANGE

Stephanie A. Schmitt

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Public Policy.

Chapel Hill
2013

Approved by:

Maryann P. Feldman

Howard E. Aldrich

Richard N.L. Andrews

David D. Dill

Lynn Williford

ABSTRACT

STEPHANIE A. SCHMITT: Assessment and Rankings Efforts: The Effect on Institutional and Program-Level Change
(Under the direction of Maryann P. Feldman)

This dissertation considers the effects of national rankings, specifically the 2010 National Research Council (NRC)'s Data-Based Assessment of Research-Doctorate Programs, on higher education institutions' behavior. National research studies of graduate education require significant resources and data, yet it is uncertain how universities make substantive, transformational changes based on participation in such quality rankings studies or their results. The dissertation provides quantitative survey results complemented by qualitative case studies to describe responses and various institutional changes that occurred as a result of the NRC study.

Evolutionary change and higher education assessment theories support the assertion that incremental changes occur most commonly within institutions regardless of the external pressures from quality rankings studies. This dissertation shows that quality studies such as the NRC can influence decision-making and improvement efforts when universities embark on change processes under serious deliberation with strong leadership and appropriate support resources. The degree of learning and organizational change depends on the perceived validity of the

study, feasibility of accompanying data collection and analysis processes, and underlying value and use of the study results.

This research will be significant for university administrators, the general public, and policymakers. University leaders and decision-makers can engage in efforts to see how peer institutions treat the rankings and engage in improvement opportunities. They can also determine whether their own institutions could manage large assessment efforts in more effective manners. Policymakers will be interested in the results because if all the funds, time, and effort spent on rankings projects result in minimal substantive action on campuses, they may wish to revamp the projects to make them more amenable to continuous improvement processes. Private market implications for obtaining necessary research and student data are discussed as ways to meet public and governmental demands for accountability, assessment, and quality control of higher education.

This work will contribute to the body of knowledge about rankings and assessment studies, particularly reviewing how they serve as information instruments to influence change and decision-making. This dissertation hopefully provides insight into policy tools, institutional structures, and processes to contribute to long-standing improvement in doctoral education in the United States.

ACKNOWLEDGEMENTS

This dissertation and my degree would not have been possible without the support and guidance from my family, colleagues, classmates, and friends, including my committee, Professors Maryann Feldman, David Dill, Howard Aldrich, Lynn Williford, Pete Andrews (and previously, Gary Henry). Each of you brought your own expertise and special interest to my project, for which I am incredibly grateful. A simple thank you is not sufficient for your support and most especially, your time. My promise is to “pay it forward” in assisting other graduate students and improving graduate education into the future through this research and my chosen profession.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAU	Association of American Universities
AIR	Association of Institutional Research
APLU	Association of Public and Land-grant Universities
CGS	Council of Graduate Schools
IPEDS	Integrated Postsecondary Education Data System
NORC	National Opinion Research Center at the University of Chicago
NRC	National Research Council
NRC Convocation	NRC Convocation on Analytic Uses and Future Directions
NRC study	National Research Council's Data-Based Assessment of Research-Doctorate Programs in the United States
Rankings studies	Used synonymously with ratings, league tables, and score or report cards as methods for assessing the quality of higher education institutions

CHAPTER 1 – INTRODUCTION

Assessments of higher education in the United States began in the early 1900s but have taken on greater prominence in the past thirty years as studies transitioned to publications that were more accessible by the general public (Altbach, 2010 and 2012; Hazelkorn, 2012). Graduate education was the initial focus for ranking studies, but now they also cover all aspects of undergraduate and professional education. The *US News & World Report* magazine first published its rankings in 1983, which coincided with the first reputational assessment of research doctorate programs by the National Research Council (NRC) (Brooks, 2005). *Money* magazine's value rankings premiered in 1989. These are just a few of the most popular examples purporting to rank the quality of aspects of American higher education institutions, including undergraduate programs, individual departments or graduate disciplines, and associated costs of attending a university.

Rankings studies are useful as they allow various constituents to make more informed choices. Audiences use information from the ranking studies in different ways, some better than others (Kuh, 2009; Berrett, 2012; Hazelkorn, 2007, 2008, 2009, 2012; Altbach, 2012; van Vught and Ziegele, 2012). The general public, especially prospective students and their families, welcome any insights into the attributes of the multitude of programs available to them. Prospective faculty and employers also review the statistics and other factors that feed into the rankings.

Funding or grant agencies could judge prospective applicants and proposals based on their rankings and associated data.

This dissertation begins with this broad view of the role of ranking studies and specifically looks at the reported and possible uses of the most recent NRC study. The focus is on extrapolating the NRC study as a form of external pressure and reflecting on resulting change processes, or the lack thereof, on university campuses from this information instrument. This study is a multi-level evolutional analysis with an emphasis on the institutional forces that can affect change processes.

While there are known issues of reliability and validity with the NRC study methods, this dissertation will not be a critique of the various methodologies or assumptions that led to the results. Rather, it is intent on reviewing the use of the study and its results as information instruments and its associated impacts for institutional change. Not all forms of response to rankings studies are necessarily appropriate. Additional focus on the use of results from an evolutionary, institutional perspective will provide greater context for the effectiveness of such studies within the assessment landscape for overall higher education quality improvement efforts.

Research Questions and Significance

This dissertation reviews how rankings studies have influenced organizational change by institutions of higher education, especially at the graduate education level. As a result of the increased public scrutiny and cross-institution comparisons, ranking studies have been used as policy instruments leading to action, including graduate program evaluations, policy changes, recruitment and public relations activities, and resource allocations. *How do institutions respond to rankings and*

quality assessments? How do institutions define, monitor, and improve quality assessment? How is the overall quality landscape for institutions of higher education affected? Specifically, how do universities themselves use the NRC study as a policy tool for enacting changes on their campuses? Institutions of higher education can respond to rankings studies in either productive or dysfunctional ways, both of which can influence quality assessment practices.

Myriad factors influence change decisions at universities. It is difficult to attribute changes caused solely by one study or assessment project, because universities are complex organizations operating in multifaceted environments. When considering what elements and dimensions might go into a comprehensive account of change in graduate education, multiple interdependent factors come to mind. For example, institutions could make changes either to influence their graduate programs' rankings or alternatively, to improve their campus offerings after the release of poor rankings. Funding levels, program growth, historical contexts, external pressures, policies related to enrollment, tuition and teaching/research assistants, and other graduate education factors such as faculty hiring and grant administration could be part of a comprehensive analysis and discussion. Only through original data collection and detailed case study is it possible to begin to understand the central questions posed in this dissertation.

This work acknowledges that audiences use the ranking studies in different ways, described above with student, parent, employer, university faculty, staff and administrator, policymaker, and general public reactions and responses. The

dissertation focuses only on use of the assessment results from an evolutionary, institutional perspective at both the university and graduate program level.

The body of literature reviewed in Chapter 2 informs this discussion. Survey outcomes and case studies demonstrate opinions and tangible examples of institutional change resulting from the most recent edition of the NRC study. The degree of learning and organizational change depends in large part on the perceived validity and utility of the instrument and study. Thus, the surveys described in Chapter 3, with results presented in Chapter 4, and case studies presented in Chapter 5 provide insights. While the dissertation does not attempt to address causality between rankings studies and organizational change, the motivating context and policy implications for this research are highlighted in the next section.

Motivating Context and Policy Implications

Universities are under many pressures concerning quality, affordability, and accountability, so their reactions to information about their performance, both publicly and internally, are important to understand. More and better information is expected to lead to stronger decision-making and thus better outcomes for organizations. The primary purpose of this research has been to determine how and whether assessment data and rankings studies, such as the NRC, have been used to provide information for policy choices and decisions leading to action by institutions of higher education.

Global Perspective

Broadly, higher education is becoming increasingly homogenized as the competition intensifies for recruiting and retaining the best students. Global

university rankings are increasing in scope, quality, and importance, often generating as much attention among the top tier research campuses as in-country studies. Strategic university partners crossing national boundaries – for alliances, shared research or academic initiatives, and faculty and student exchanges – are driven by a desire to be aligned with other highly-ranked institutions, even if only at the reputational level.

Within graduate education, the educational models and expectations in the United States, especially doctoral-level training programs, are spreading to other countries. Governmental oversight, accreditation practices, centralized graduate school structures, and university cultural differences play a role. Competition to decrease "brain drain" from within the country and to attract the diversity brought from an influx of foreign students of varying nationalities is leading to revisions among faculty and university administrators worldwide. This line of research and structural change encompasses a different body of literature, review of governmental and professional oversight structures, and organizational culture considerations than those addressed in this dissertation. However, the global context for trends in graduate education is important to note as the backdrop for the importance of studying university decision-making and responses to rankings and assessment studies.

The NRC study is only one recent example of a prominent, national assessment, quality, and rankings study. This dissertation reviews the context of university response to external pressures through the lens of the NRC study. While the findings are not generalizable to all rankings studies, domestically or globally,

this research will provide insights, best practices, and recommendations for how such studies can be used to implement change and program improvements on campuses.

Public Policy Considerations

Further analysis of the NRC study and results is significant to policymakers, university administrators, and the general public. Federal and state public resources, not to mention the universities themselves, support these studies, including the NRC. Particularly when one considers the diffused effect of resources and staff time spent collecting and analyzing the input data, impacts on a campus can be substantial not only for the central university but also for every academic department or program required to collect and assess information. This work will contribute to the body of knowledge about rankings and assessment studies, particularly reviewing how they influence change and decision-making on campuses. Using a multi-level survey, the research also contributes to better understanding of alignment on goals and change processes between the central administration and its component graduate programs within the university.

Policymakers will be interested in the results. If all the funds, time, and effort spent on rankings projects result in minimal substantive action, the projects should be revamped. Governments and other policy decision-makers may also provide support for implementing best practices to facilitate action. It is also possible that at minimum the federal government's role in directly funding such efforts could be reconsidered. Additionally, policymakers and study owners may recognize

necessary methodological changes in the rankings projects to make them more amenable to continuous improvement processes.

There are implications for the private market as well given the increasing demand for data and analysis on universities from varying corners (Glenn, June 2010 and December 2010; Wiley, 2009; Suskie, 2010; Lederman, 2012; Dill, 2011; Shavelson, 2010). Research is beginning to focus more closely on measures for comparison across higher education institutions, such as the Gates Foundation's Context for Success project that focuses on outcomes measures for education (Fain, 2012). Private companies, such as Academic Analytics, Thomson Reuters, and Elsevier, have sprung up to meet some of these data demands. They provide data reporting and analysis, for a fee, to university campuses with a primary comparative focus on faculty research and productivity metrics. They cater to institutions who want more detailed information about themselves and also appeal to the competitive nature of rankings and elite peer status (Thomson Reuters, March 2013). Broader and more accessible metrics are still needed. The for-pay aspect of these data may be cost prohibitive for some universities and raises questions about the independent nature of these firms. Yet the emergence of these innovative companies show a business model exists to meet university demand for quality data.

Not only do universities wish to study and make comparisons about their peers, but there are also increasing demands for accountability, assessment, and quality control of higher education. These forces come from all levels of government and funding agencies as well as the general public. Assessments need to be well-

constructed to present accurate information that is useful to addressing these concerns now and into the future.

University leaders and decision-makers will benefit by understanding how peer, competing institutions use the rankings. They will also find it useful to determine whether their own institutions can manage large assessment endeavors in more effective manners. Such efforts can influence competitive positions for universities, affecting student and faculty recruitment, access, external research support, and tuition levels.

Against the backdrop of the NRC study, this dissertation aims to review how universities are using the NRC study data and results as policy tools for continuous improvement and change on campuses. The NRC study has historically been seen as gold standard research on graduate education deserving of wide dissemination (Lederman, 2005; Hicks, 2008). Together with the resource and use considerations described above, all these reasons justify additional research as to the effectiveness of the project in creating change and improvement in doctoral education.

National Research Council's *Data-Based Assessment of Research-Doctorate Programs in the United States*

The National Research Council is part of the National Academy of Sciences, a private, nonprofit institution that provides science, technology, and health policy advice under a congressional charter signed by President Lincoln in 1863. It provides services and advice to the federal government, the public, and the science and engineering community. The NRC is the main operating body of the National Academies, and its researchers and experts provide data, analysis, and guidance on

many current policy issues and decisions faced by the nation (National Academies' website).

One of its primary contributions to the higher education policy arena has been the decennial assessment of the quality of research doctoral programs in the United States. The NRC study has traditionally called upon administrators, institutional researchers, faculty, and staff to gather large amounts of data and respond to various survey requests. The National Academies augments the collected data with data on faculty publications, citations, and awards. The staff then analyzes the results and releases a thorough data analysis and assessment, including rankings, of the quality of doctoral programs nationally.

This comprehensive research and assessment project has occurred three times. The NRC study was first conducted in 1983-84 and again in 1995-96. The most recent study was conducted in 2006-07 and was named the *Data-Based Assessment of Research-Doctorate Programs in the United States*. It used 2005-06 as the baseline year for data collection purposes. The results were released in the fall of 2010 and included a written volume plus a massive Excel spreadsheet of program-level data about participating doctoral programs on multiple NRC study variables. This dissertation focuses on this latest iteration of the study, which is described in detail along with the context and controversies surrounding its release.

NRC Current Release and Errors

The most recent study's stated goal was to provide faculty, students, and policymakers with an in-depth look at the quality of those programs that produce our future researchers, teachers, and practitioners (NRC study project website; NRC

study report and database). Policymakers, university administrators, and academic scholars in the higher education and assessment research arena appear to accept this goal by disseminating, highlighting, and building on the NRC's methodology and practices in their own work. While the NRC's methodology and findings have been targeted areas for academic research, analysis, and criticism, minimal work has occurred reviewing the institutional effects and changes resulting from a large research effort at the graduate level such as this.

The results released in fall 2010 included characteristics and ranges of rankings using the NRC study methodology for over 5,000 programs in 62 fields at 212 institutions. A revised NRC study methodology was also released at this same time, which included an overview of the two ranges of rankings, the S and the R rankings, and descriptions of the 20 key variables that contributed to them (NRC study project website; Jaschik, May 2010). The resulting multiple ranges of rankings, in lieu of an ordinal list of discrete rankings of programs, was an intentional step on the part of the NRC study commission to show the complexities and inherent inaccuracies in basic assessments of graduate education (NRC release webinar, 2010; Kuh, 2009; Lederman, 2005). Simple rankings of an endeavor as complex as graduate education cannot take into account the full breadth of training opportunities, research foci, scientific methods, and disciplinary standards in use in graduate programs across the nation.

Errors in the NRC data were immediately recognized. In the week period between the embargoed release and the full release, the NRC asked institutions to report known errors immediately. Basic examples included one university that had

been labeled as a private rather than a public institution, plus the entire field of Computer Science had issues with a key variable on student outcomes (Computing Research Association, 2010). The rankings were re-run for Computer Science, and the public release on September 28, 2010, produced a different set of data. The NRC then asked institutions to report all found errors by November 1, after which they would determine whether there were issues substantial enough to re-run the statistical methods again to achieve another set of ranges of rankings or if they would simply publish the errors on a public website for users to see.

On April 21, 2011, the NRC released another version of the Excel spreadsheet that included updates as submitted to the NRC and revised ranges of rankings. The NRC reported it had received queries about approximately 450 doctoral programs from 34 institutions. This revised spreadsheet was immediately found to include new technical errors in time to degree and completion data for programs in the history of art, architecture, and archaeology field and a revised spreadsheet was posted a week later. The revised spreadsheet incorporated some technical and factual corrections from the original release and noted four key areas where reviews were requested (see Appendix 1.1). The April 28, 2011 version is the final dataset posted for use by universities and students and in any research projects desiring data on doctoral programs.

The revised ranges of rankings did not address all issues raised by institutions or accommodate all requested changes. Specifically, much of the background data behind several of the reported variables, such as faculty publications data, was not released. The NRC did not accommodate any requests

from universities to revise their faculty lists as originally reported in 2006 (Glenn, March 2011). Though not stated directly, several of the areas where requests were not addressed included key components to the rankings that, if changed, may have significantly influenced the outcomes. Once the rankings were released, the NRC did not entertain requests that might have been perceived as gaming the results and rankings. One's assessment of where this boundary line existed is acknowledged as open to interpretation.

Errors and misunderstanding are centered primarily on the faculty measures, specifically concerning the count of faculty who were included and their omitted publications and citations (Glenn, June 2010). The NRC has not provided information publicly about the process used to gather publications data for faculty other than to say their efforts mined Thomson Reuters (ISI) Web of Science, a multidisciplinary publication and citation database product. Public discussion at the 2011 NRC Convocation, discussed in greater detail below, and on websites and project email listservs suggests that when universities or graduate programs mined for their own faculty, they typically found much higher publication counts than reported by the NRC, even with the embedded error taken into account.

Similarly, the faculty allocation process in the NRC study is a point of marked confusion (AAU Association of Graduate Schools letter, January 2011; Glenn, December 2010; Drahl, 2010). Institutions were asked to provide faculty lists for each participating program using faculty commitment to the doctoral program, as defined by advising/mentoring and teaching service. If an individual was named as core faculty in more than one program, their workload and productivity was split

evenly. Affiliated faculty members were allocated across programs that named them through calculations based on dissertation committee service. These steps were taken to ensure no one individual achieved more than one hundred percent effort across all their affiliated doctoral programs, even for individuals with joint or multiple departmental or program affiliations. This method clearly creates confusion when multiple programs claim someone.

Once the results were released and people saw the outcomes of these allocations, questions and concerns arose. Faculty in the programs did not understand why someone they considered a core faculty member only partially counted in their graduate program, both for headcount and workload purposes. The argument arose that the allocation method harmed highly-interdisciplinary faculty and programs. Such interdisciplinarity is an already-accepted and growing tenet of strong graduate education (AAU Association of Graduate Schools letter, January 2011; Glenn, December 2010). The differences in faculty productivity mentioned above were not explained simply by the faculty lists or allocation processes.

The second key area where errors and confusion occurred centered on how the NRC study assessed funding for students. There was a complex funding grid in the program questionnaire[1] that identified various sources of funding, as well as multiple combinations of those sources of funding. According to the NRC methodology guide, the study only used a couple of these line items to describe student funding. Most notably, the NRC results reported the percentage of students with research and teaching assistantships, but they did not include the combination

[1] Specifically question E8 in the NRC Program Questionnaire.

lines that included these sources of funding along with other sources. So for example, a student who was funded solely on a TA would be counted as a funded student in his program's calculations; yet the student funded by both a TA and an external scholarship would not be counted in her program's funded variable. This method of counting reduces the overall percentages of funding shown for students and under-reports programs' true emphasis on student funding. This type of analysis is to the disadvantage of public institutions that might have more combinations of funding for graduate students than private universities. Several public universities mentioned this issue as raised by their programs at national conferences such as the NRC Convocation or publicly on email listserv groups.

Key Characteristics, Findings, and Critiques

Even with its shortfalls, the NRC study results could be viewed in any number of meaningful ways. There are some key takeaways from the report and ranges of rankings of fields. Important and valued variables in most fields turned out to be publications per allocated faculty, citations per allocated faculty, faculty awards, and percent faculty with grants. Inherent in these outcomes is the allocation of faculty across programs as described above, which dictated who was included in these findings and at what percentage of effort.

For publications, the NRC used Thomson Reuters to conduct searches, which will necessarily miss some faculty, even when trying to match on zip code or name and university combinations. Some fields do not publish or value (only) journals that are tracked by Thomson Reuters, and so some fields will have important publications left out. This topic was discussed at the NRC Convocation, and the

NRC said these issues were consistent and should not disadvantage one university or program over another. Given time and resources, they did “the best they could”. Yet some disciplines cried foul about their perceived low outcomes on these measures of faculty productivity, especially given their importance in the weightings and outcomes (Cole, 2011; American Sociological Association, 2011; American Mathematical Society, Mucha, 2011; Jaschik, March and April 2011; Computing Research Association, 2010). Because the NRC did not release these data publicly, the reported values for publications and citations cannot be recreated, thus questions of accuracy and misrepresentation have remained.

When reporting the percent of faculty with grants, the NRC used the respondents to the faculty survey as the denominator and not the full faculty count for a program, a decision that could be subject to unintended bias. This variable also does not capture the amount of grants, just the percent of responders who reported grants, again leading to potential misrepresentation. For example, a small school with each faculty respondent holding a $1k grant may look stronger on this characteristic than a research-intensive university with half the faculty respondents holding multiple grants of six to seven figure dollar amounts. At the extreme, if a university program had only one respondent and she happened to hold a grant, their program would be reported at 100% compared to another university that had high response rates with a mix of grant holders.

There is some reported concern that the diversity and gender variables are negatively correlated with quality in the NRC study methodology (NRC release webinar, 2010; NRC Convocation, 2011). Thus, programs that do well on these

measures will in fact see their overall ranges of rankings harmed by the weights and coefficients associated with the measures. The NRC addressed these outcomes by indicating that diversity is labeled as important among respondents and at the same time indicated these measures will not carry a great amount of weight unless other quality characteristics are equal. They also emphasized in the report, and at the release press conference and NRC Convocation, that the weights are not as important as the absolute values when looking at each variable. Thus a program that performs well on their gender balance among faculty or students, for example, can see how well they do compared against their peers even if the rankings themselves do not take this variable into account due to the methodology employed. Even with this explanation, the programs that sensed their rankings were reduced because of perceived strong performance on diversity measures are questioning the validity of such an outcome, especially one that is now publicly presented to researchers and prospective students (American Mathematical Society, Mucha, 2011; Drahl, 2010).

The NRC study imputed missing data on the 20 key variables reported in the spreadsheet. In many cases the imputed value is the average value for the whole field, a rather crude mechanism for imputing data, especially since it does not appear to have been consistently applied. In some cases, missing data were reported as zeros, whereas in others, the missing data were given the assigned average value. This choice can significantly impact a program's faring in the ranges of rankings. For example, a zero assigned to percent of first-year students with full support weighs heavily downward for a program compared to assigning the average

value for the field to missing values. The NRC did correct this specific value in the final April 2011 release to allow programs with no first-year students in a given year to remain as missing data instead of any imputed value.

Program Differences

The NRC study results also showed some distinct differences among programs they included. Several disciplines were not included in the study because of uncertainties in how best to quantify quality measures, especially in more professionally-oriented doctoral training. Fields such as Education and Social Work were excluded even though as disciplines, their faculty would argue that their PhD programs are research-oriented. While the NRC tried to incorporate emerging areas of study, interdisciplinary fields created issues. The more-established disciplines appeared to fare better with the NRC taxonomy and methodology.

Applying the Biglan model can help explain why it is so difficult to generate taxonomies for studies such as the NRC. Through his work surveying faculty at two higher education institutions, Anthony Biglan (1973) provided an approach for classifying different academic areas. The model offers various ways to structure and understand academic disciplines, including continua along three dimensions: hard-soft, pure-applied, and life-nonlife. To compare across dimensions and disciplines is unwise, Biglan argues, as the characteristics and social constructs within a discipline have an impact on the type of output a program produces. The NRC study can be viewed through the lens of aspects of this conceptual framework, especially due to the authors' goal not to aggregate and compare rankings across disciplines or at the

institutional level. Structuring a taxonomy along these dimensions to ensure appropriate comparability would be difficult, yet more meaningful in the long term.

In the most recent iteration of the NRC study, two field areas are especially noteworthy within this framework. First, Communication Studies as a field is very broad for two main reasons. It is a relatively new field and so there is not as much agreement among the faculty who would have been surveyed as to what characteristics are most important to quality. The NRC field includes both Communication Studies and Journalism, which are very divergent fields at some schools. Journalism will cover print and electronic media, public relations, and advertising. Communication Studies can include theater and performance based studies as well as organizational behavior, rhetoric, and communication theory fields. These varying characteristics of this one NRC field make it very difficult to perform valid comparisons across the programs and institutions. Campuses could self-select peers among the whole field and compare themselves on the individual data variables, yet the ranges of rankings would not be terribly meaningful in this scenario. They cannot be re-run while only factoring in certain programs within a whole field.

Second, there is also wide variance among the weights and rankings in the Pharmacology and Toxicology NRC field. The S and R rankings vary so widely that some institutions see their programs perform very well under one methodology but look relatively weak using the other methodology. The key variables mentioned above (e.g., publications per allocated faculty, citations per allocated faculty, faculty awards, and percent faculty with grants) carry great importance in the R method for

this field but are more weakly influenced in the S method, which provides one explanation for the wide variance between the two methods. Many of the programs included in this field are interdisciplinary and engage in research not only with support from but in partnership with the National Institutes of Health and National Institute of Environmental Health Sciences. Faculty often hold joint or adjunct appointments between the Institutes and the partnering university. The faculty data submitted as part of the NRC study may be skewed as a result, which could impact their faculty productivity outcomes. Because of their different funding streams and employer expectations, individuals who are engaged in these doctoral programs may not publish at the same rates, carry their own grants, or even be affiliated directly with the listing institution. Thus, their productivity may appear less than that which is traditionally expected of faculty for valid reasons due to work left uncaptured in the NRC study.

Although there are now known issues with the NRC study, its release was highly anticipated. This chapter next turns to this time period and provides some immediate reactions to use of the study results.

Media Coverage and Usage

The mainstream media did not cover the NRC study release very heavily. The study was very complex with few headlines that would be easily consumable by the average reader. Because of the lack of discrete rankings, no one could legitimately tout their program as number one among the nation. Most campuses reported the results accurately by sharing ranges of rankings without an attempt to provide an ordinal list of rankings. The media releases and university websites

reviewed were transparent about the process and results (Jaschik, September 2010; Inside Higher Ed; Boston University website; University of Michigan website; Cornell University website; Duke University website; University of California-Los Angeles website; University of California-Berkeley website; Boston University website; University of Virginia website). Most universities took the upper end of one of the ranges of rankings and extrapolated their success from there, typically suggesting their university had a certain number of programs that could have fallen within the top ten percent or quartile nationally in quality. Few universities appeared to make university-wide counts, listings, or statistical claims. All of these responses were within the spirit of the NRC study as described in the National Academies' release news conference and media press release.

For several weeks after the release, blogs and articles did appear on discipline-specific or general higher education publications such as the *Chronicle of Higher Education* and *Inside Higher Ed*, among others. In large part, these articles questioned the NRC results and methods. Faculty were suspicious of the results, especially from the S ranking method that showed quality indicators in a bottom up fashion based on surveys of faculty, which was a new methodological approach for the community to understand. Historically the NRC study maintained a traditional rankings approach with large reputational components based on peer opinion surveys, which are generally frowned upon as statistically illegitimate quality measures. The 2006 NRC study commission undertook efforts to change its methodology to address these criticisms (Jaschik, May 2010). And while the higher education community welcomed this change, in part due to the distrust for standard

rankings efforts, the complexities and unusual reporting for the NRC results still led to confusion about the processes employed once released. This distrust even led to some calls for reintroducing some elements of reputational assessment in quality studies (Glenn, September 2010; NRC Convocation).

Statements on the immediate use of the NRC data and results were varied but primarily indicated mining the data for relevant information to each campus and program. Many of the deans, chancellors, and presidents who released statements or provided information on websites upon the release talked about needing more time to process all the data and to determine how they were going to use the study to improve the quality of their graduate programs (CGS Statement, 2010; Jaschik, September 2010; Brown, 2011; various university websites and press releases as captured by the author).

Discipline-specific attention similarly included a desire to understand better the data variables and not necessarily focus on the rankings. Those programs that did well wanted to see where they were strong compared to their peers so they could tout that message to prospective students and their deans. Other programs wanted to know where they did not do well so they could focus improvements on those key areas, presuming the NRC study will occur again. While some programs that did not fare as well questioned the validity of some of the variables, they were able to move past the rankings and look at the specific variables for areas of improvement (Glenn, December 2010; Drahl, 2010; Russel, Gibeling, and Weiss, March 2011; Grasgreen, 2011; Brown, 2011).

The NRC study and its results appear to have elevated the discussion of graduate education on campuses, including some of the key data variables that were highlighted (NRC Convocation, 2011; Glenn, March 2011; Russel, Gibeling, and Weiss, March 2011). Though campuses reacted to the study release, it was not immediately apparent that campuses were using the NRC results right away for making decisions about the future. This finding at the time of the release was expected due to the complexity of the study and the abundance of data that were released.

In the NRC study release webinar and press materials, as well as via remarks at the NRC Convocation, the study commission was clear in its inability to endorse the ranges of rankings and stressed they were illustrative only. They continued to emphasize that the data were the most important outcome of their study. Their hope was that universities would use the data to compare themselves on individual factors. It was not their goal to produce an ordinal list of quality, and they admitted to intentionally making it difficult to massage the data and results into such a list. Commission members expressed a desire for campuses to use the data to further the conversation about the importance of graduate education and allow students and faculty to select the key variables that are important to them. They can then focus on how to make changes to improve those key variables.

Other uses described at the release time involved other audiences. Students may want to look at one key variable, such as how long it will take them to complete a degree depending on where they go. Faculty or administrators may want to look at how much growth in diversity measures they need to achieve to be more

competitive with their selected peer group. The NRC and its commission would support these uses of the data well beyond any use to rank programs and make decisions based solely on rankings. Herein lies the issue of focus for this dissertation.

The following section describes the NRC Convocation in more detail to lay the foundation for a review of uses of the study results.

NRC Convocation

On March 4, 2011, the National Academies hosted the *NRC Convocation on Analytic Uses and Future Directions* in Washington, DC. NRC staff, NRC study committee members, senior leaders in higher education, institutional researchers, higher education policy and nonprofit representatives, graduate education leaders, and the media gathered to discuss the current and planned usage of the NRC study and its results. Opening remarks set the stage for ensuring that the universities – the bearers of most costs and the recipients of most benefits from the NRC study – had the key voice in the use and dissemination of the results. Universities were also asked to lead the national discussion about the future direction of these rankings and assessment efforts.

As this conference occurred prior to the release of the revised database, several presenters and audience members spent time addressing the errors uncovered in the data with cautions for how best to use them. The overall theme from most universities represented was that the NRC study shed light on core characteristics of doctoral education in the United States and provided more openness nationally about assessing graduate education efforts.

Several key outcomes and themes arose over the course of the day. There were calls for simplifying the data collection and methodology in the future. If universities and graduate programs can focus on fewer variables with standardized definitions, the NRC study outcomes will be more action-oriented, understandable, and usable by the participants and audiences. Specific areas of use were the focus of many presentations and comments and included managing and improving programs, strengthening the campus program review process, identifying competitive positions, providing consumer information, and some public relations usage such as with governing or legislative bodies.

Campuses indicated across the board that to really use the NRC data, programs must cull it down to individual characteristics. The study as a whole provides too much information. To be useful and actionable, a campus or program must pick one key characteristic, select program peers, assess program standing among the peers and national means, and set goals based on these results. Working toward these goals can then influence future perceptions of quality level and overall rankings. The data were very important. With the data in hand, campuses can initiate conversations about program quality and continuous improvement among faculty, administrators, and program chairs. Prospective students are increasingly interested in such metrics and comparisons as well.

Several presenters took these ideas a step further and shared which tangible areas for further review and action were used on their campuses. These focus areas included admissions data, completion and time to degree data, diversity measures for students and faculty, funding levels, and faculty productivity measures.

SUNY-Stony Brook, for example, shared the process their campus used for the NRC data. They set a level or threshold of performance for their programs relative to their peers, such as top 25 or top 25%. Programs that fell below this threshold would then have a focused review with very tangible outcomes. It is possible then that programs falling below the threshold could be scaled back or bolstered to bring them into the higher tiers based on campus priorities.

Other campuses described similar studies and uses of the NRC study data. Most stressed that the NRC data and peer rankings should be but one input into this process. Campuswide strategic planning processes should include additional data from campus and other national sources such as the Integrated Postsecondary Education Data System (IPEDS) and the Council of Graduate Schools (CGS). In general, such inputs and quality measures help to position resources and determine where to invest in future directions. Additionally, with proper study, quality data and peer comparisons can assist in determining which factors might be high impact areas where change could most contribute to program improvement.

Some conclusions can be drawn from the NRC Convocation and university presentations on data use. Not only can program quality studies lead to specific decisions for change, they can also be a catalyst for internal conversations about the availability and use of program-level data on the campus. When national studies such as the NRC bring program quality into the realm of discussion, it helps campuses understand appropriate measures of quality, whether and how to collect the data on their specific campus, and metrics for supporting programs. If collected routinely, campuses can then benchmark themselves against their own progress,

regardless of the NRC comparative data or even if the specific study is conducted again.

Future Efforts

Concluding comments by NRC leaders reflected on the variety of uses for program improvement that could result from their study. Ralph Cicerone, Chairman of the National Research Council, stressed that research doctoral programs are a national asset impacting society. Higher education is under public demand for accountability and for producing quality programs, and the NRC study was seen as meeting these demands. Dr. Cicerone stated his belief that the value of the project stems from the fact that the universities contributed so much and added intellectual energy into the project. Similarly, Bill Colglazier, NRC Executive Officer, said that the "value in the exercise was the database itself". Even with the issues the study encountered, the comparability of the data helps to shape the conversation about graduate education among faculty and leadership in higher education nationally.

It was unclear whether the National Academies would pursue another iteration of the NRC study. In fact, leaders implied that the NRC could have a convening role in data collection and validation but not take the lead on the study going forward. The sense was that most of the future directions for quality studies and assessment data collection and validation will be up to the higher education community, and specifically the graduate education community. Leadership and ownership for these efforts need to have groundswell support instead of directives from the NRC.

Speakers pointed out the expense of waiting too long before beginning the next phase of the study. It was suggested that perhaps even collecting these data every two years would be both more useful and less expensive in the long run than a decennial study. Many university leaders agreed that continuous improvement is more relevant and impactful than one massive study. Then campuses could review trends over time and determine whether actions and decisions are having the intended effects on quality improvement.

Several speakers at the NRC Convocation also stressed the need to consider unintended consequences from the study results. The results could drive behavior, perhaps in ways the graduate education community may not want to see, especially in regards to interdisciplinary programs. First, as described above, highly interdisciplinary programs generally might not fare well in a strict taxonomic study like the most recent NRC assessment. Because there are a variety of fields represented, peer comparisons are difficult. The NRC study does not easily allow crossover comparisons between fields for the rankings data. While unlikely, if this outcome leads institutions to forego fostering more interdisciplinary programs, graduate education and scientific research will face significant setbacks at the very time when research bridging fields is seen as most innovative and successful to addressing the key issues in today's world.

A second unintended consequence relates to future iterations of the study. The potential for misuse of the data in a second run is more likely unless the methodology changes significantly or stricter definitions of variables are identified. At the time of data collection and submission in 2006, the NRC methodology was not

certain. Thus there was not as much potential for manipulation of the data, as institutions did not necessarily realize how the NRC was going to perform certain calculations. Without a change, universities may submit data in ways they think will lead to stronger rankings and outcomes in the next study.

A third key element relates to the faculty listings. While the NRC was clear on how faculty were to be included and listed, now that campuses can see how this total or allocated number of faculty were used in a number of variables, especially as it relates to faculty publications and citations, institutions may be more prone to select faculty more intentionally in the next iteration of the study. For example, it may be that clinical or adjunct faculty who had served as dissertation chairs or committee members were included on faculty lists – in accordance with the NRC guidelines. These faculty members may not be actively engaged in other aspects of the graduate program, including publishing or producing significant amounts of research. Many of them are affiliated with campus or nearby research facilities, as described above in the Pharmacology and Toxicology discussion. It may be tempting for the program to leave them off of the next round of faculty listings for these reasons, when in fact they did meet the criteria NRC established for including faculty. Understandably, administrators and program chairs are questioning why someone who is not necessarily expected to contribute in some research areas could then be included in program variables on those same characteristics and ultimately harm their affiliated programs.

In short, there appeared to be general consensus that if the NRC study were to be repeated, many of the issues brought up at the Convocation will need to be

addressed, including the need to establish firmer definitions for faculty and data elements, a simpler methodology, more transparent variables and data, and outcomes that are geared toward more usable and action-oriented decisions (NRC Convocation, 2011; Grasgreen, 2011).

Outline of the Dissertation

This dissertation began with a discussion about the policy relevance of this line of inquiry and background on the NRC study. Chapter 2 provides an assessment of the relevant evolutionary change and higher education assessment literatures and gives an overview of theoretical frameworks focused on education rankings. Universities, with their layers of decision-making and decentralized structures, serve as strong examples to study and determine the rationale behind actions.

Chapters 3 and 4 then describe and provide results from surveys of university administrators and graduate program faculty directors regarding their opinions and use of the NRC study. Chapter 5 provides the basis and results of studying these ideas via three university case studies. Many facets of university life can impact how universities respond to rankings and data about their own performance. The continuum of change and its influences are highly individualized and best informed through in-depth study of specific universities and their graduate programs, even though results are not generalizable to all campuses.

The dissertation concludes in Chapter 6 with a discussion about the policy implications and best practices in this arena and possible areas for additional research.

CHAPTER 2 – THEORETICAL PERSPECTIVES

Few studies focus on rankings assessment as a form of external pressure on universities and the critical dimension of the resulting change processes, or the lack thereof, on university campuses. Three key thematic perspectives are relevant, each of which relates to the research questions identified in Chapter 1.

The first area of substantive literature details assessment in higher education. This section includes broad overviews of assessment and informative evaluation. A discussion of the literature on external influences for assessment, such as the government and accountability calls from the general public, is provided. An additional element of assessment includes theories on the study of use, primarily emphasized through literature sharing the rationale behind the adoption and use of research.

Next, there is a literature on rankings and institutional quality approaches that relates to methodological issues in the field and calls to standardize the processes and analyses. This literature is much more tangibly rooted in research design and description than theoretical but nonetheless provides insights into what factors may make the NRC study more meaningful for implementing process improvements. Ranking studies are generally organized around several themes, such as reputational assessments, faculty research productivity measures, and student-oriented analyses. Academic literature critiques each approach. Research centers

on different units of analysis, ranging from the institution down to the departmental or academic program level. Rankings research is broken into three categories: a historical look at rankings of higher education institutions; the stability of rankings over time and what factors influence this phenomenon; and the uniqueness of quality indicators at each institution and the difficulty this presents for wide-scale comparisons among universities.

The third area of this review is the broad field of institutional change. Literature highlights the usefulness of assessment rankings based on how effective they may be at leading to institutional action. Given that institutional change theory is such a vast field, work that can be readily applied to the university setting is the focal point. The literature is discussed in two sections, one focused on organizational behavior and change in general, and then a second section more focused on institutional change as it relates to quality and rankings studies.

The most significant works relied on are Gormley and Weimer's (1999) conceptual framework for organizational response where they analyze best practices surrounding assessment report cards across several sectors, including higher education institutions; the study by Feldman, et al. (2002) on the adoption of use of a particular innovation or technology transfer strategy; Aldrich and Ruef's work (2006) on sustaining continuous change and assessment and quality practices; and Dill and Soo's work (1999, 2005, 2006, 2007, 2010, and 2011) on the potentials for enacting change, functional or otherwise, based on quality and rankings studies. These frameworks, theories, and positions are described in greater detail below.

Throughout this chapter particular attention is paid to work focusing on graduate education. The chapter concludes with a brief review and discussion about methodological techniques and issues of influence, which aided in the selection of appropriate methods for this dissertation research. The theories and frameworks described in this chapter informed the development of the surveys and case studies discussed in later chapters.

Higher Education Assessment

The first two sections of this Chapter provide literature and theory related to the research question, *How do institutions define, monitor, and improve quality assessment?* Assessment practices are growing in the higher education realm with focused attention on external pressures, accountability, and overall improvement practices for universities, both academically and for student support services.

External Influences

The concept of assessment generally will be highlighted, especially as a policy tool for social betterment. Societal gains result if critical, timely, and informing evaluation and assessments are performed (Mark, Henry, and Julnes, 1999; Henry, 2000; Henry, 2003; Henry and Mark, 2003). Evaluation processes evolve through three stages: determining the common good, selecting a course of action, and adapting the course of action. Emphasizing the “paradox of persuasion”, or the idea that evaluation must balance support of broad social goals yet not be the ultimate goal lest the study risks losing credibility (Henry, 2000), provides a framework for structuring process change effectively.

Both direct and indirect influences impact the decisions and practices of organizations. The ways in which priorities and change processes at universities are influenced by external factors, such as the NRC study and subsequent focus on graduate education nationally. Organizations aware of impending evaluations, or rankings and quality studies, will implement actions and make decisions driven by appropriate and potentially inappropriate responses.

Research focused on general higher education assessment provides insight into why universities may in fact implement changes as a result of rankings studies. One of the most critical elements surrounding action and the assessment of higher education is the heavy involvement by the federal government (Vaughn, 2002). Federal and state governments fund major portions of the higher education enterprise in the United States through student financial aid and the support of research endeavors via grants, fellowships, and training opportunities. Even in difficult budget times, this support remains in place, though pressures do continue to mount at both the state and federal levels. In some corners governmental support for higher education is increasing, at least in the balance between federal research support and state budgetary support for public institutions.

Increasing pressures then follow for institutions to justify their use of funds through accountability measures. Access to higher education has expanded in large part due to an appreciation for its impact on the broader economy and societal progression. No longer will the government allow universities to function completely autonomously – not when the future hinges on the success and proliferation of higher education. At minimum baseline responses to key accountability, and

increasingly affordability, standards are expected of universities (Srikanthan and Dalyrymple, 2007; Massy, 2013).

While the government and public's desire to evaluate educational programs and spending may be understandable, care must be taken to ensure the government can assure quality of educational initiatives and institutions without inserting itself too broadly into long-standing academic affairs values (Vaughn, 2002; Wellman, 2003). Accreditation practices and indirect measures of quality through data collection and dissemination are becoming increasingly popular in calls for reform and openness. In a challenge to the traditional role of higher education, Alexander (2000) states, "Once it has been established that the primary purpose of higher education is to serve the economy, then it becomes the responsibility of the state to ensure that the institution is held accountable in successfully achieving this task" (page 427). While an acknowledgment of public accountability is appropriate, higher education as an institution must also retain its core principles to educate students broadly and engage in cooperative relationships with the public and governments. Researchers call on the higher education community to become more actively engaged in identifying and assessing quality indicators lest risk losing all authority to rankings studies, valid or not (Hazelkorn, 2013).

Agency theory can be used as a key framework for examining the university-government relationship (Kivisto, 2007 and 2008). Information asymmetry and goal conflicts are commonplace in higher education and fall nicely into agency theory modeling. These attributes can lead to principal-agent problems, especially when the government attempts to oversee the quality of the university. Theory would

predict that governments as principals do not trust universities as agents because they can act opportunistically – shirking, pursuing only prestige, distorting data – if they are not held accountable for quality and resources. Thus governments have developed output-based governance tools to measure quality, focusing on metrics like graduation rates, first placement data, and debt levels to look at higher education in a cost-benefit lens. Also, governmental regulation and intervention ensues (Kivisto and Holtta, 2008). At the undergraduate level, the division between the purchaser and consumer of higher education can increase information asymmetry (i.e., parents who pay their children's tuition). It could be argued that at the graduate level, this becomes even more complicated – both because some graduate and professional students are payers and consumers, and at the same time some graduate students receive support from the universities themselves.

Principal-agent theory could also be used to view how quality ratings and information asymmetries can impact knowledge about a campus and its educational offerings (Kivisto and Holtta, 2008). Some inputs assume that high-quality research is correlated with robust graduate programs or higher doctoral student demand. Yet conflicting information about output quality and the possibility of gaming the rankings process can lead to different interpretations of measurement information. But even if more information is publicly available, questions still exist about how to weight the information and ensure the comprehensiveness, validity, and reliability of the findings. More efforts are needed to understand these trade-offs and improve the data available to the universities, which in turn becomes the basis for their accountability responses to public and governmental entities.

Beyond issues of funding, such factors are also key in the increased push for higher education accountability and measurement of student learning outcomes by government and policymakers (Shavelson, 2010). Universities are increasingly being asked to establish goals on quality and performance, provide the support and resources to assess those goals, and engage in feedback loops to assure improvement over time. While continuous evaluation of student learning outcomes for Shavelson is more of a focus for accountability and accreditation practices, lessons can be learned regarding overall quality enhancements and learning on campuses. In particular, both lines of inquiry require institutional leadership. They also become more effective with centralized structures and resources to support data collection and analysis. These types of practices become institutionalized over time, embedding themselves in the culture and expectations of the campus community.

Similarly, an openness to change and faculty engagement are critical both for student assessment as well as program and institutional change processes. Graduate education has only recently started to deal head-on with accreditation and program review emphases on student learning outcomes. Even more so than undergraduate learning, graduate education has historically been faced with long-standing perceptions about successful outcomes, many of which are being called into question as US competitiveness appears to weaken and traditional job markets for graduate alumni become less stable. Campus processes to assess these changes, evolve training programs, and improve student preparation for a variety of

careers – each having heavy professorial involvement – are beginning to infuse graduate education nationally.

Although it does not include an analysis of the United States, a project by the Organisation for Economic Co-operation and Development (Santiago, et al., 2008) reviews the importance of higher education in a knowledge-based global economy and offers suggestions on the most effective methods for assessing quality. This form of competition is the final external influence to be discussed. Increasing calls for cross-national and cross-study comparisons have the potential to lead to best practices and lessons for implementing change processes and improving quality studies (Merisotis, 2002; Lorden and Martin, 2000; Dill and Soo, 2005; Ewell, 2004). Quality assurance, including formal accreditation and overall evaluative activities, varies across countries on many factors, many of which can be compared to analyze strengths. Practices will depend on the focus and level of review and whether the study is intended to assist with improving an institution or program or is more centered on formal responses to accreditation and accountability expectations (Santiago, et al., 2008).

Other differences include the role of the government, participation of key stakeholder groups, the timing of reviews, public perceptions, and the rewards or sanctions associated with a given country's processes. Each of these factors lead to a variety of circumstances depending on the study or country's practices under review. Competition across them can lead to the establishment of common best practices.

Institutional behaviors may change due to the existence of measurement practices. Assessment will be most successful when external measurement goals and techniques align with internal institutional goals and performance indicators. Quality is necessarily nebulous, with rankings often limited in their focus or scope in any methodology. Yet these pressures and competitiveness across institutions to have the highest-quality higher education opportunities can influence behaviors. The rankings efforts in use in higher education today can be improved. Clusters of rankings and more accurate and broad data about education will provide better information for use in sustaining improvement practices. The NRC study is a recent example of an attempt to improve these types of analyses. The study's methodology, while complicated, has the potential to bring more elements of consistency and reliability to this area of study.

Use

Research focused on use and the utilization of research provides some foundational bases for studying why and how universities could enact changes and improvement efforts as a result of rankings studies.

Much of the seminal work in this area comes from Weiss's work (1977, 1979) describing multiple models of the utilization of research and how policymaking and governmental decision-making can be enhanced through the use of research. Researchers have responsibility to ensure their work is understandable and usable by policymakers to enable informed decisions and contributions to the policy landscape. Many would say the NRC study failed in this area because it was so time-consuming and complex. Rather, if research study coordinators want the policy

implications of their efforts to bear meaning, considerations in presenting findings and making recommendations include political factors, finding champions or owners of key policy themes, and goal setting and identification.

The primary order of research is to begin with defining a social problem, identify missing knowledge, acquire social research and data, interpret the problem solution, and ultimately lead to policy choices (Weiss, 1977, page 12). This linear model, however, can be limited in practice where a more diffuse and overarching pattern of engagement can be more successful in enacting change. At universities seeking to assess and interpret quality, the policy implications may already be at the forefront of the effort. Instead, a mode of assessment that is conducted for intentional use (Patton, 2000) may be more effective. If the end goals and stakeholder needs are taken into consideration as the study is being generated and framed, then there is a higher propensity for longer-term adoption and use. The role of the researcher, Patton argues, is to serve as a negotiator between these needs and to bring the scientific rigor and expertise to the study. In this manner, the study is more adaptive and meets the validity and credibility thresholds for appropriate social science research. Yet it also meets the end goals of use and has the potential to serve as a change initiator.

As an example, Van Dooren (2005) ran a study in Belgium to determine what conditions must be in force for organizations to engage in performance measurement practices. The use of data was deemed critical for information-driven policy and management tools. Market pressures were influenced by data as well, including league tables. Both adoption and implementation of performance

measurement were influenced by the scale and resources of the organization, the measurability of the services offered by the organization, and the link between goals and measured indicators.

Within the context of the NRC study, these models would be most applicable if the study is reframed to focus more closely on the end goals of the users – the programs, faculty, and students – instead of the long-standing focus on rankings and a drive for prestige. These attempts were made in the most recent iteration but the other issues that plagued the study did not lead to a recognition of this change in focus. The universities and program faculty likely needed to be more engaged in the study goals and setting common definitions to ensure adoption once the results were released. Additional discussion of these themes will be addressed in the sections and chapters to follow.

Quality and Rankings

Much of the literature on assessing institutional quality relates to methodological issues in the field and calls to standardize the processes and analyses. This section continues with a focus on the research question, *How do institutions define, monitor, and improve quality assessment?* Ranking studies are generally organized around several themes, such as reputational assessments, faculty research productivity measures, and student-oriented analyses. Academic literature critiques each approach. Research centers on different units of analysis, ranging from the institution down to the departmental or academic program level. Many studies focus on a certain discipline or thematic area (e.g., Political Science

departments in Dometrius, et al., 1998 and Miller, et al., 1996, or Nutrition programs in Greger, 2006) instead of a broader taxonomy of all relevant areas of study.

As described above in the external influences section, quality studies are also conducted to meet the call for accountability in higher education. Accountability can be viewed as a triangle balancing state or governmental priorities, academic concerns, and market forces (Burke, 2001). External accountability and internal improvement practices can be achieved through a variety of means, including process audits (Massy, 2001 and 2013), performance measurement (Volkwein and Grunig, 2001), and governmental or accreditation checks (Richardson and Smalling, 2001; Wiley, 2009). These activities are common in graduate education through the external review process for both grants and program quality in general. Most existing quality studies, in particular the ones that result in rankings, place too much emphasis on reputation, a call made in most research on higher education quality rankings.

If there were greater involvement among universities and faculty to improve the system of quality assurance and assessment, policies would likely lead to greater external accountability to both governmental entities a well as the general public. Such practices also have the potential to influence institutional behaviors, a theme discussed in greater detail in the following sections. These accountability pressures are not waning (Gumport, 2005), especially at the graduate education level where vast amounts of governmental support contribute to the research enterprise (Gumport, 2005; Buela-Casal, et al., 2007). The interplay between doctoral education and training, academic research and grants, and US federal

government support is in effect at all research-intensive universities, especially those that are primarily doctoral-serving institutions that participated in the NRC study.

Thus it is increasingly important that reform occurs in how rankings are developed (van Vught and Ziedele, 2012). Only then will universities be able to move beyond the traditional rankings toward emerging and more beneficial areas of use. Conveners of rankings studies can reflect on the various stakeholders who should have continuous engagement during the development and implementation of rankings. Openness and transparency can lead to multidimensional metrics, including program-level inputs that measure qualities such as research, teaching, and engagement. These broad arrays of data have the potential to develop efforts and studies that take different missions, audiences, and uses into account (van Vught and Westerheijden, 2012).

As an example, van Vught and his colleagues review various classification systems in effect and critique their methodologies in order to arrive at a set of best practices for their own rankings method, the U-Multirank system. Still early in its public use and data collection phases, the U-Multirank system may already be encountering some of the difficulties that befall most rankings projects, such as lack of understanding and university distrust (O'Leary, 2013; Hazelkorn, 2012). Yet these efforts to expand the scale and scope of traditional rankings methodologies are necessary to move the discussion forward.

This section continues by providing an overview of various aspects of quality and rankings literature. For ease in discussion, the research is broken into three categories: a historical look at rankings of higher education institutions; the stability

of rankings over time and what factors influence this phenomenon; and the uniqueness of quality indicators at each institution and the difficulty this presents for wide-scale comparisons among universities.

History

Several researchers have compiled historical looks at higher education assessment and quality studies and shifts over time (Hattendorf, 1986; Tan, 1986; Brooks, 2005). Quality indicators, methodologies in use, weighting factors, and definitions of key metrics are key areas of attention in any research on this subject. Institutions and scholars can be pushed to develop new, combined measures of multidimensional quality. Single rankings based solely on a faculty citation count, or a reputational ranking coming from a select group of faculty or administrators, cannot possibly capture the full range of quality found among university graduate programs, especially with different constituencies and competing priorities involved. The NRC study was one such effort that attempted to combine multiple methods and metrics and avoid resulting ordinal rankings.

Research also focuses on evaluations of the rankings studies themselves (Van Dyke, 2005; Dill, 2006; Dill and Soo, 2005, 2007, 2010). Some studies are valid and deemed high quality by researchers, the public, and universities alike. But more often issues with the studies are detailed, including the measures in use and the pursuant outcomes for chasing prestige based on these rankings. Overall university rankings do not appear to help guide student choice to more individualized programs, and in fact league tables can encourage institutions to pursue prestige-

based changes that reflect well in the rankings, an outcome Dill deems a zero-sum game (2006).

Lessons can be learned from these efforts even as attempts are made to improve upon them. By providing overviews of the quality studies in a landscape fashion, it becomes even more apparent that these studies grew up in an unproductive manner by building on simple metrics and outcomes. The historical evolution path has only marginally changed due to these calls for caution against dysfunctional responses to the studies.

Also, as is common in this literature, undergraduate studies with their more quantifiable metrics are attempted and assessed more frequently than quality studies of unique graduate education programs. The lessons from the undergraduate education attempts can be applied as continued growth at the graduate level is undertaken.

Several examples of studies that were closely reviewed are worthy to note. National performance indicator efforts in Great Britain were put in place in the 1990s to assess higher education and led to universities undertaking their own performance reviews (Cave, et al., 1997). The authors discuss who can use the data, at what level it is generated (i.e., program or institution), the government's scope in assessing higher education institutions (i.e., past assessments or forward thinking only), and the value placed on the process (i.e., how valid and reliable it is). Case studies such as this assist in determining best practices and methods for change processes using assessment data as a foundational basis.

Similarly, a review of reputational studies that ultimately led to the creation of a set of principles of quality and best practices in university rankings (IREG) focused on the Center for Higher Education Development (CHE) rankings in Germany (Soo and Dill, 2007). This ranking study is carefully designed and generally accepted as very strong, valid, and lacking many of the controversial elements of other rankings studies due to its balance between reputational components (including faculty and student opinion) and data-driven analysis. The study also results in groupings of universities and programs as opposed to ordinal rankings as outcomes. Yet even stronger studies can be deemed to have concerns, including the accuracy of the data, the dysfunctional effects of university rankings on universities' behaviors, and self-reported data elements – each of which were found in Soo and Dill's analysis. The NRC is the most comparable study in the US higher education realm, especially due to this study's decision to present recent results in groupings and data in unweighted forms. Soo and Dill highlight the "infotainment value" of the rankings, an acknowledgment that this type of information is relevant and in demand, yet an understanding that more work must be done to provide better data that has potential to contribute to true academic quality assurance.

Stability and Inertia

Many researchers have pointed to the stability in rankings over time as a sign that significant, sustainable changes cannot occur at universities as a result of the ranking studies. The opposite might be expected if market forces were dictating student and faculty choice based on actions universities undertake to improve their standings (Eccles, 2002). Academic inertia is partially to blame for this outcome. At

the same time, lag effects are widespread and the methodologies used to assess quality do not easily allow for new changes or trends to appear in the overall findings (Vaughn, 2002; Usher, 2013; Selingo, 2013). Especially when reputational factors are part of a study, the common methodologies undertaken and in use today do not allow for significant shifts over time. If senior administrators or faculty are expected to rank a discipline's programs, their attitudes and perceptions take substantial time or a significant event to alter.

Hazelkorn (2012) likens the global university rankings to the Olympics and cautions that on the surface the winners and losers can be viewed too simplistically. When one looks more deeply at the rankings and the inputs, some of the smaller universities may in fact be more successful and efficient in their outcomes based on the resources and number of faculty available to them. The stability found at the top of the top of the rankings may be misleading if research endeavors to find true change in university priorities and quality improvements. Such improvements may be in force but without attention to that level of detail, it may be missed.

Another reason the top rankings may not evolve much over time relates to the existence of halo effects, or the fact that one graduate program within an institution may be thought of so highly that other programs are perceived more highly than may be deserved (Dill, 2006). Strengths at the university level, especially in reputation-heavy studies, can bleed over into the programs even if they alone are not at the top of their fields. This outcome strengthens the need to involve additional broad factors in these studies to ensure as accurate a picture as possible is developed.

This finding also holds true when comparing factors that influence the rankings of graduate programs compared to undergraduate education. For example, in one study undergraduate indicators included in reputational rankings, such as average SAT scores or high school rank, explained a majority of the total variance in reputational rankings for doctoral programs as well, plus total departmental size and research levels were highly correlated with NRC rankings (Grunig, 1997). Studies such as these in part led to the various methodological choices undertaken in the revisions for the most recent iteration of the NRC study described in Chapter 1.

Research activity and funding levels, in particular, can be included in the halo effect phenomena. Higher-rated universities have been found to enjoy more success in increasing research funding allocations (Geiger and Feller, 1995; Sine, Shane, and Di Gregorio, 2003; and Moy, et al., 2000). Funding is becoming more concentrated in institutions within the top tiers, and researchers are questioning whether or not this outcome is advancing research and policy appropriately. The research and faculty at institutions or programs who happen to not be at the top of a simplistic ranking methodology generally ought to be able to compete just as successfully for research and training grants. Sine, Shane, and Di Gregiorio's (2003) findings, in particular, show that general prestige based on the *US News and World Reports* rankings influence licensing rates of university inventions and technology. Once prestige effects become embedded within a discipline or field, then they can be difficult to overcome. In part these are the examples that lead to increased attention on studying how prestige impacts institutional change.

Similar to halo effects, the anchoring effect phenomenon – making judgments based on limited, available information, such as an institutional ranking, regardless of the validity of that information – has been shown to influence longitudinal institutional rankings (Schmidt, 2010). In this example, *The Times Higher Education Supplement's* World University Rankings were used as a natural experiment for measuring anchoring effects to show that perceived quality of specific fields led to the reputational rating for whole institutions. These types of studies are further evidence for the exclusion of reputational factors in assessment studies. Such ordinal rankings drive reputation in a continuous cycle and are not a valid measure over time of program or institutional quality.

And finally, a common finding is that high multicollinearity exists among variables that potentially influence institutional quality, thereby making it difficult to determine the relationship between final rankings and the actual causes behind them (Toutkoushian, Dundar, and Becker, 1998). Burris (2004), in particular, discusses such findings in an analysis of the academic prestige hierarchies relevant at the hiring stage for graduates, which ultimately shows a lack of change regardless of attempts at inter-institutional hirings and collaborations among high and low status institutions. The perceived status of a university is maintained through a variety of factors, such as faculty hiring and recruitment goals, even in the face of intentional efforts to restructure priorities. Even more efforts could be put into place to attempt to make these shifts more amenable to change over time.

Each of the methodological issues described in this section needs to be taken into consideration when trying to understand the potential for using assessment and

quality studies to improve graduate programs. The stability of rankings over time and the lack of perceived change remain factors to overcome. If the stability and inertia can be explained by some of these factors, the graduate education community will not necessarily lean on the rankings as a key measure of quality. Rather, looking deeper into the key metrics for graduate program quality and distancing these metrics from some of the issues like halo and anchoring effects has the potential to improve the types of data available for use in this area.

Unique Interpretations of Factors

Many rankings studies rely on universities self-reporting data in response to surveys or queries about their programs. This practice can lead to misinterpretation of the questions and therefore biased data and results. Much of the data is strongly dependent on the unique characteristics and administrative structures within a university (Eccles, 2002; Hattendorf, 1996). Having minimal external, independent validation of reported data could enable attempts by institutions to influence their results through inaccurate or inappropriate reporting (Dill, 2006). These potential outcomes, intentional or otherwise, hinder the overall validity of the rankings studies as presently configured. Standardized reporting methods and common definitions of key metrics and inputs at universities are critical to have in place, yet are currently lacking especially at the graduate education level.

Several researchers have reviewed the national landscape on report cards on educational performance measures in attempts to address the discrepancies in definitions of performance indicators (Breneman, 2005; Ewell, 2004; Goldstein and Spiegelhalter, 1996). Surveys and policy instruments such as those from the

National Center for Public Policy and Higher Education and the National Survey of Student Engagement take quantitative data elements and review them as assessment tools. Commercial rankings are suspect because they do not truly represent institutional performance and are generally consumer-driven instead of driven by accurate and measurable data. Especially appealing are any attempts to standardize the data, format, and report structures to allow for better comparisons across studies and universities. As part of the push for greater accountability, governments could advance such statistics and efforts to link teaching and learning outcomes with performance evaluation for universities. For this to be successful for graduate education, a need is present for program-level data to exist in easily accessible and comparable formats. This situation is not readily the case today, but efforts have begun and will continue, a theme returned to in later chapters.

Data on institutional factors that relate to program quality are often in the eye of the beholder. Tuition and fees, a seemingly simple example, are often reported differently within educational frameworks – across state systems or between public and private institutions. Care must be taken to understand the complexities behind a set of variables lest some misuses of data become possible (Berrett, 2012; Altbach, 2012). Policymakers and governmental entities attempting to understand costs and benefits associated with funding higher education should be especially cautious in understanding the nuances behind a set of data and rankings (Berrett, 2012). Institutions answer queries based on their definitions of the variables on their own campuses, regardless of how their peers may interpret and answer the same questions. While measures to control for construct validity may be present in each

data collection effort, the fact remains that the final results may be biased until common definitions are adopted, a difficult undertaking.

Next, the third section of this chapter will drill even deeper to review literature that focuses on institutional change and highlights the usefulness of quality and assessment rankings based on how effective they may be at leading to institutional action.

Institutional Change Theory

A vast literature exists on organizational change and institutional change dimensions. This last section of this Chapter provides literature and theory related to the research question, *How do institutions respond to rankings and quality assessments?* It also explores the larger research question, *How is the overall quality landscape for institutions of higher education affected?* Of specific focus are works on institutional change in higher education settings or broader theories that can be readily applied to the university setting. One discussion below is focused on organizational behavior and change in general, and then a second section is more focused on institutional change as it relates to quality and rankings studies.

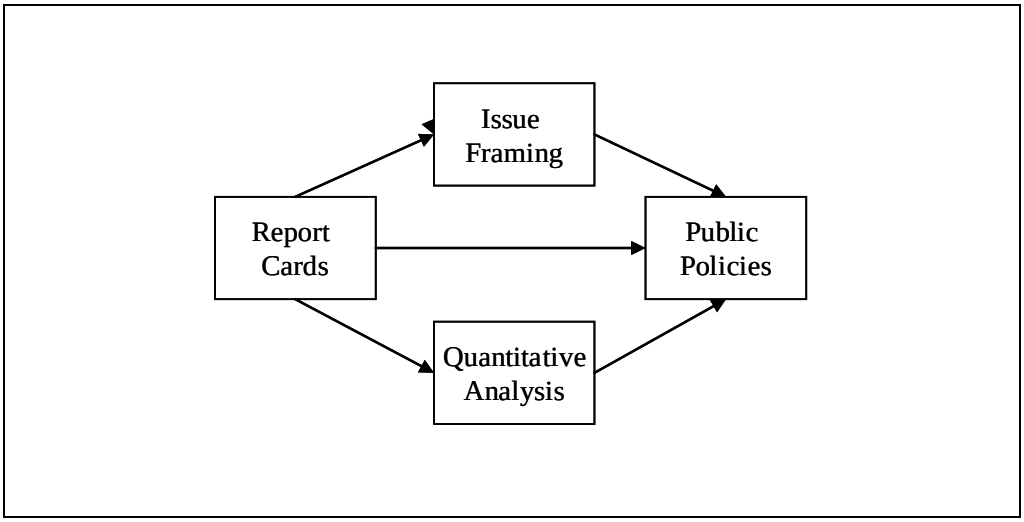
Organizational Behavior and Change

Although cynical views about organizational response and propensity to change are widespread, many theories provide some support and predictive ideas for how institutions may make changes in response to external pressures, such as quality assessment and rankings efforts and data.

The most significant work relied on as a theoretical model for this dissertation is from William Gormley and David Weimer, a book titled *Organizational Report*

Cards (1999). They analyze best practices surrounding assessment report cards across several sectors, including higher education institutions and healthcare entities. Report cards and assessment data are increasing in importance as policy tools, especially via efforts that emphasize valid and understandable indicators of performance. Depending on the study specifications, report cards assessing the quality of organizations or other entities can be meaningful to policymakers in their decision-making processes. Such data, if well-constructed and disseminated appropriately, can also garner the public's interest and focus debate on key performance indicators.

Figure 2.1 – Modeling Flow of Information and Decision-Making from Report Cards



-Page 115

Figure 2.1 above shows the flow of understanding and predicts potential impacts from report cards. Political, academic, and technical perspectives each play a role in this model, particularly in the development stage for an assessment study with report cards as expected outcomes. The planned end users for a study, from

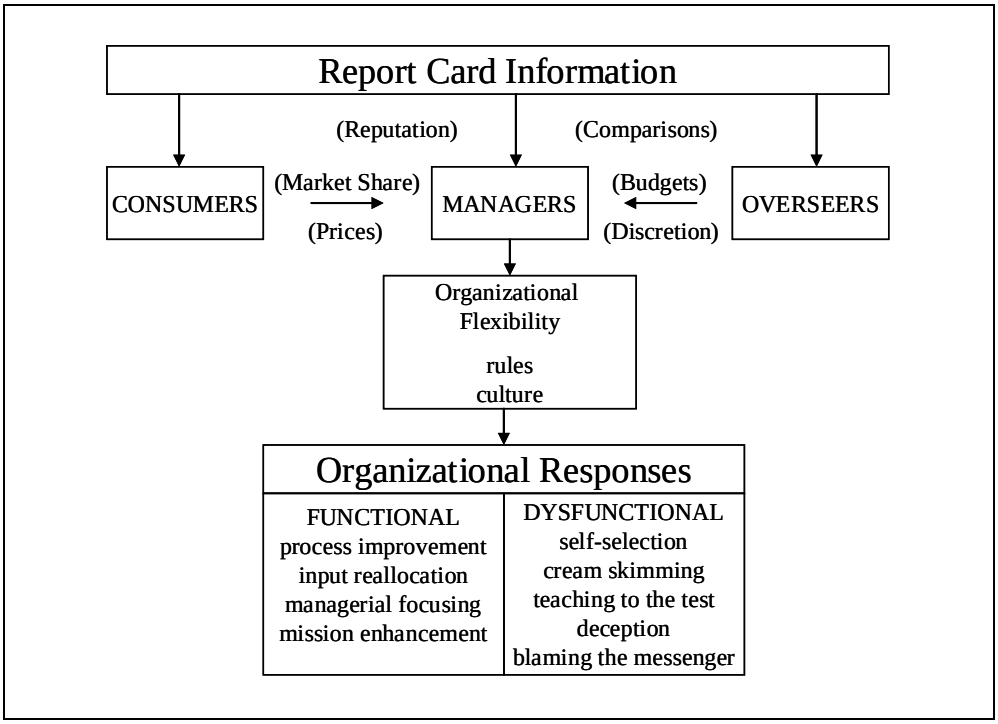
the general public to sophisticated policymakers and their staff, can appreciate varying levels of political realities and quantitative mechanics. This in turn may influence the study's design, included metrics, and weighting mechanisms that lead to the creation of the report cards and rankings.

Both paths have the potential to lead to policy decisions. The data from report cards can be used in quantitative analyses to provide detailed information for policymakers. The level of sophistication in a particular study's methodology both introduces validity of the rankings outcomes and creates complexities in end user acceptance and understanding of the outcomes. Similarly, the rankings can be one input when framing policy issues in the general public and with researchers. Mixed levels of report card and rankings data can be provided to either frame a policy issue as desired or to allow end users to evaluate the results themselves and arrive at their own conclusions for action.

When considering how this model may be applied to graduate education specifically, multi-faceted use also becomes a factor. Central administrators and faculty, both program directors and chairs as well as rank faculty, have varying levels of interest and understanding for certain quantitative methods that factor into a quality study's outcomes. The NRC study, for example, perhaps went too far in their sophistication of methods; the study became indecipherable by many expected end users thus rendering the results more suspect. Yet the data can be present for those administrators and faculty willing to use them to support campus research initiatives and improvements in graduate programs.

Similarly, graduate education is ripe for framing analyses, primarily because many external constituents do not have a full appreciation for graduate training and what it entails. Advocacy efforts at both the state and federal levels by graduate educators and administrators often need to begin with the basics of what master's and doctoral education involves. The ability to offer simple metrics and data on quality would prove helpful in justifying the expenses and broadening an understanding of the value of graduate education. These themes described here guided the development of the multi-level administrator and graduate program surveys used in this dissertation and described further in the chapters to follow.

Figure 2.2 – Conceptual Framework for Organizational Response



-Page 135

Gormley and Weimer's conceptual framework for organizational response is depicted in Figure 2.2 above. It shows not only the varying roles relevant to and

impacted by organizational report cards, but also the possible uses and responses to the rankings. There is potential for multiple feedback loops. As different roles use data from rankings and report cards, their responses can vary within the bounds of organizational structure. The overseers and managers roles in the model maintain the most discretion for implementing changes, such as faculty and staff leadership on campuses. Graduate program directors have oversight for their curriculum and faculty serving to mentor students. Both levels of use of assessment data and rankings can influence organizational response.

The central importance of organizational culture is depicted in this model as well. An institution's openness to change will influence many of the actions taken after the release of rankings, for example. Functional responses should create an environment for continuous process improvement and overall enhancement of the academic mission of the institution or program reviewed. It has the potential, within the confines of organizational culture and priorities, to lead to sustainable and beneficial change over time. Dysfunctional responses, on the other hand, potentially create a hostile environment where changes could occur solely to influence future strong rankings. For example, universities may decide to participate in studies that would only rank their strongest programs, or they may choose to allocate resources to the determinants of final rankings in some study, to the detriment of other needed upgrades or enhancements on campus.

Institutional Learning

A secondary conceptual model that proves useful is adapting the Feldman, et al. study (2002) on the adoption of a particular innovation or technology transfer

strategy, specifically the use of equity agreements. They argue that the adoption of equity use is in response to problems encountered with traditional licensing agreements and adoption strategies vary based on layers of institutional learning. Institutions will change and adapt based on learning, the diffusion of best practices, and organizational incentives.

Changes in institutional behavior also occur based on increasing accountability pressures (Dill, 1999). With a focus on improving teaching and learning capabilities, universities need to become learning organizations through key organizational structures and best practices. Case studies aid in reflecting on the importance of faculty oversight, universities support structures for evaluative and continuous improvement activities, and the importance of peer evaluations and comparisons. The culture at a university can drive change, especially if that culture embraces innovation and values open assessment and improvement processes.

While not directly related to rankings and quality studies, these themes can be consistently applied to questions of institutional change and adaptation in that context. Institutional learning, both from within and among their peer group, played a large role in the adoption of equity, and can be similarly applied to organizational learning from comparative rankings studies. Past performance together with practices gleaned from peers and collaborators allow universities to be more entrepreneurial and receptive to new ideas. Quality and rankings data can be key inputs into institutional learning processes and peer comparison activities.

Sustaining continuous assessment and quality processes in institutions of higher education is a more difficult task. Boyce (2003) presents an overview of the

literature in this area. She conducts a meta-analysis on organizational learning literature specific to change in institutions of higher education, theorizing which factors influence the success and sustainability of change. Key factors include leadership, structural changes, levels of innovation, and an organization's level of commitment to the change process. Without a combination of these factors, institutions revert to former processes and practices without actually moving forward. These findings are similar to the theories Aldrich and Ruef explore in their book *Organizations Evolving*; transformational change rarely happens, for even if the volition is present to do so, the environment must be ripe to sustain it over the longer term. The forces of variation, selection, and retention must be present and interact before transformative change could result.

Cohen, March, and Olsen's (1972) seminal work on organizational choice and decision-making reflects a view of higher education as a form of organized anarchy. Universities have problematic preferences, unclear technology, and fluid levels of participation. Their garbage can model highlights the possibility of evolution toward lack of oversight and constant changes in decision-making processes, together with decisions that do not necessarily resolve problems. Similarly, reviewing institutional change processes during times of challenge and contraction on campuses can be meaningful opportunities for identifying change factors (Rutherford, et al., 1985). Structural, social, and personal factors – each hinging on political processes – promote or inhibit change. Any type of radical change must first occur at the core value level on a campus before procedures or processes can sustain long-term change. Yet this level of change does not happen easily or often.

Peer Effects

These themes surface in additional research agendas as well that show long-term sustainable change is difficult and that external pressures and competition drive institutional behaviors. Case studies and quantitative survey analysis have been used to gauge university management perceptions of identity and image (Gioia and Thomas, 1996) and to monitor how institutional behaviors change as a result of being ranked and evaluated by outside agencies (Espeland and Sauder, 2007). Other work studies the effects of rankings on student support services, including the effects of higher-rated institutions offering more subsidies to their students to compete with one another for top quality students, disadvantaging public institutions (Winston, 2001).

Thus, perceptions of a university's prestige and standing among its peers can catalyze the change process. Of particular importance is a university's envisioned identity and image among its peers. The values and core tenets related to an institution's culture and identity will need to be in line with these goals before leadership pushes toward any type of sustainable change that may impact peer standing. The goal of achieving a top-10 ranking can become paramount, even to the detriment and losing sight of other goals on a campus. Given the ranking studies' role in measuring these standings, aspirational images may be defined and perpetuated through participation in the studies or by virtue of the results themselves.

Organizational effectiveness and environmental effects are intertwined and influenced by external perceptions and judgments of an organization (Pfeffer and

Salancik, 1978). Rankings exist as one central driver of external perceptions of program quality. As stated, “information, regardless of its actual validity, comes to take on an importance and meaning just because of its collection and availability” (page 14). Many rankings studies meet these criteria. The process of collecting the data, awaiting the results, the media attention provided to the results, and the consequent stability and prevalence of the rankings all point to the increasing relevance of studies regardless if they are accurate or producing meaningful improvements on campuses.

Another related theory of change highlighted in DiMaggio and Powell’s work (1983) posits that organizations mainly change through processes that make them more similar to each other. Homogeneity in structure, culture, and outputs result from attempts to deal with uncertainty and organizational constraints. The concept of competitive versus institutional isomorphism, particularly the mimetic processes, helps explain institutional reactions to rankings studies. It appears difficult to completely separate competitive forces from purely institutional forces. Universities compare themselves to peer institutions in a variety of ways, to the point that imitation is a factor in many facets of higher education decision-making. Organizations model themselves after successful peers as they attempt to formulate and zero in on their own institutional goals. Again, assessment studies have processes and outcomes that lend themselves to this model of institutional behavior. Change may occur, though more easily in the form of modeling peer behavior instead of the more difficult substantive, core changes in institutional identities and values.

Change Management Processes

Change processes on a campus can be managed in a number of different ways, often dependent on the culture at a particular organization. In his more theoretical work, Clark (2003) argues a steady state of incremental changes eventually leads to a new status quo. "Cumulative change rolls a university forward", he claims (page 112). Volition to change precedes substantive change, and because not every institution is willing or able to enact core changes, inertia could more easily become the guiding hand.

In a case study review of an in-depth change management process at one Austrian university, Meister-Scheytt and Scheytt (2005) show that even a rationally-planned change process encounters roadblocks and paradoxical behavior on the part of key players. Each subsystem within a university culture operates with its own logics, and if not managed effectively, such behavior can often lead to detrimental effects in overall university change progressions.

In translating this idea to the NRC study, because of its emphasis on individual graduate programs, as opposed to the entirety of a university's graduate education agenda, the study could be subject to such divergent agendas. Even if a university chooses to implement a strategic planning process using the NRC data and outcomes as inputs, individual departments may have different priorities or subvert the common goals to their own benefit. The rankings and assessment data mean something different to different audiences, including whole disciplines as well as individualized units within them.

Yet true organizational transformation is rare (Aldrich and Ruef, 2006). Inertia and the status quo serve as the predominant state of affairs for organizations. Changes in goals, boundaries, and activities must each occur. As changes occur, they must be repeated or reproduced so that ultimately the new routines become embedded in organizational culture. Otherwise, the organization will not retain the changes. External forces can often lead to internal variations and are an important driving force toward organizational evolution.

Applying this tension between inertia and external forces of change to universities, and specifically to the effects of rankings studies, can be telling. Universities are large and increasingly complex organizations, especially when factoring in the resources, faculty expertise, and resources necessary to maintain a critical mass of graduate training programs. Goal identification and movements toward enhancing quality, as part of an en masse effort, would be exceedingly difficult to manage. Thus, the external influence from rankings studies has great potential to impact institutions of higher education and their reluctant change management processes. So while not common, or possibly even expected, change brought about by rankings and peer effects can occur.

Taking external pressures a step further, the historical context for transformations also maintains significance in organizational change (Aldrich and Ruef, 2006; Pfeffer and Salancik, 1978). Evolutionary changes occur over the life course of an organization, and in particular, “external events interact with an organization’s own actions to drive the pace, pattern, and direction of change” (Aldrich and Ruef, 2006, page 161). Cohort effects and period effects could

differentiate between the institutions that participated in a NRC study and those that did not. Given there are only three iterations of the study to date, the cohort effect would influence institutions at different stages of their life course. For example, a long-standing land grant state university might have different reactions to the NRC results compared with an emergent technology-focused institution just beginning to grow its graduate programs. The period effect would indicate consequences that influenced all institutions similarly by virtue of having participated in a given NRC study. However, such effects should also translate at the population level and provide a common framework within which all institutions move ahead. Minimal emphasis should be given to individual, institutional changes.

Possible period effects also call into question the timing for graduate level studies. For example, the National Academies has historically conducted the NRC study once every 10 to 12 years. Other national studies rate graduate programs infrequently and inconsistently. Such in-depth studies are expensive and time-consuming, so the spaced-out timing is cost effective. A question may be whether this timeframe is relevant enough to bring about change. On one hand, because the results will be in the public realm for so long without updates, institutions have even higher incentives to look as strong as they can at the time of the study. The counterclaim is that no one may be paying attention to these reports and thus it is irrelevant to the natural change cycles and rhythms in university settings, especially budgetary cycles that have the potential to affect change processes on a campus. Regardless, waiting too long between updates and iterations is not helpful for

facilitating an environment of long-term continuous improvement as called for in this body of literature.

These findings can be translated to the higher education assessment arena by considering that institutional learning is a realistic outcome from participating in rankings studies. Institutions learn from the data collection and preparation phase, as well as when the results are released and data across peer groups are available. Universities are feeling increasing pressures to enact change, and to move quickly, unlike the historical norm in higher education where change is more consultative and long-standing (Kiley, 2012; Massy, 2013). The more robust and serious ranking studies include opportunities for sharing best practices and help identify for campuses what works well and which variables contributes to quality, at least within that particular study's definition of quality. Peers potentially learn from one another both in attempts to better their own institutions and to succeed in the competitive market for students, faculty, and resources – including near-neighbor peers as well as aspirational peer groups. Both rationales have built-in incentives for campuses to learn, evolve, and change. Especially at the graduate education level, appropriate and adequate time and data could enable focused attention on peer comparisons, process audits, and learning outcomes.

Institutional Action Related to Quality and Rankings

As indicated earlier, little academic scholarship has centered on the actual institutional actions that result from participation in rankings studies, particularly at the graduate education level, though this area is gaining interest. Research and study directors acknowledge the great care that university administrators and

researchers should take when interpreting the results of national assessment studies and conducting comparisons across institutions. Even with all the faults found in the methodologies undertaken in rankings studies, administrators should not ignore the ratings (Stigler, 1996; Lane, 1996). There is benefit in comparing results against peers. The fact that the rankings have been published is reason enough to bring focus to the campus and raise attention on any impacts, including the reputation faculty and programs have nationally.

Uses and Misuses

Hearkening back to the earlier sections in this Chapter about potential functional and dysfunctional responses for organizational change, a closer look is now taken at specific uses for rankings studies. Maher (1996), also a co-chair of the 1995 NRC study project, provided a summary of misuses and misunderstandings with the NRC data in a follow-up article to address some of the complaints that arose following that iteration's release. He argued that attempts to compare across disciplines or to aggregate key metrics to establish one's own benchmarking and rankings were inappropriate. Rather, programs should mine the data and look for areas of improvement by focusing time and energy on the methods and data. These arguments were strikingly similar to what occurred over two decades later when the next iteration was released. The most recent iteration of the NRC study embraced these suggestions and emphasized the potential positive uses of the data to delve deeper into key characteristics that matter to individual graduate disciplines and specific programs.

Several researchers explore the potential for misuses of rankings study data. Dill (2006) cautions that university responses are not necessarily positive at all times. Potentially harmful actions that universities may take in response to rankings studies include attempts to increase research funding by developing new, yet perhaps unwarranted, doctoral programs and recruiting faculty solely to influence the research inputs for rankings studies. Campuses may also increase tuition and put the additional dollars toward inputs that increase prestige in the rankings and not necessarily emphasize those issues that enhance the quality of higher education.

Brewer, Gates, and Goldman perform an exploratory analysis of the conduct of universities in their book, *In Pursuit of Prestige: Strategy and Competition in US Higher Education* (2002). They focus on four revenue sources – student enrollment, research funding, public fiscal support, and private giving – and the use of discretionary revenues. Part of their focus is on institutional objectives and how different institutions, including categories of institutions, may have differing objectives yet still be required to interact. This in turn influences their behaviors. In sum, they believe institutions with high levels of prestige can increase their discretionary revenues, which may drive behaviors. Prestige generation and the pursuit of additional prestige tends to lead toward higher research activity and emphasizing more degree programs at the doctoral level. While not always the case, their case study findings suggest that universities can follow this common approach.

Recent media reports continue to describe actions taken by universities to improve their standings in worldwide rankings, such as faculty being asked to recruit

others to participate as peer review survey recipients and the resulting backlash (Jaschik, 2013). Rankings continue to grow in interest yet the methodologies are not necessarily being improved to be more inclusive of institutional mission or to develop more inputs to capture the wide range of research being performed at universities today (Labi, 2013). The cautions from these overviews are clear. The pursuit of prestige may in fact not target improvements that benefit students or an institution's mission. Instead, pursuits to increase prestige among their peers and generate more revenues have the potential to lead to difficult choices overshadowing true curricular improvements that need to occur on campuses. Next, specific examples of studies are discussed that reviewed the use of rankings studies and their impact on university culture and future improvement plans.

Examples and Reviews of Use Studies

In their 2005 work, Dill and Soo conduct a comparative analysis of university rankings in Australia, Canada, the UK (two rankings), and the US in an attempt to assess possible improvements for higher education internationally. They ask three primary questions: is there consensus on the measure of academic quality across instruments and countries, what impact do the rankings systems have on university and academic behavior, and are there critical interests missing from the rankings? They evaluate the league tables on five counts: validity, comprehensiveness, relevance, comprehensibility, and functionality.

They find that while inputs are easier to include, they are not good indicators of quality. Outputs are more difficult to include and standardize but would eventually show a clearer picture of the value added from higher education. The five league

tables are assessed for their ability to enact action within universities, which can include means to improve quality but also ways to circumvent the instrument, data, and analyses based on manipulation of data. Government can take a key role in improving the quality of assessment activities (Dill and Soo, 2005; Dill and Beerkens, 2010). If the government intervenes, it can help ensure the same level of information is provided by all institutions and that the public interest is being served by the data and assessment activities. Without such oversight and funding, the commercially-produced assessment tools may fail to standardize the process and produce valid, reliable results and outcome measures.

These findings can be applied to this work as well. Institutions are more willing to respond to assessment of research doctoral programs such as the NRC than first-degree programs, which are more typical and easier to quantify. The study's methodology attempted to incorporate some of the quality output measures for graduate education, as recommended by Dill and Soo. It also attempted to push universities to use the data and results for program improvement rather than solely emphasize a final set of rankings. The National Academies did take the lead role, but even with that governmental stamp of authority, the study encountered issues as described in Chapter 1. Thus even using the best practices and recommendations from a survey look at rankings does not amount to guaranteed improvement in quality processes.

Several researchers have used survey tools and cases to attempt to explain how institutions might react to rankings studies. Hazelkorn (2007, 2008, 2009, and 2011) looked at the various constituents and consumers of worldwide league tables

and rankings to understand broadly how policymakers and universities themselves might be using the results. Survey findings, augmented by follow-up interviews in various countries, showed some tangible outcomes at institutions, such as university restructuring, strategic planning, and goal-setting, as well as policy-making actions such as funding allocations and institutional classifications. Even here the trend to take action with prestige-seeking in mind as the end goal, even if unintentionally, was deemed potentially damaging to campus climates. A key takeaway from her work is that global rankings are continuing to increase in importance as strategic instruments with greater potential for enacting change and affecting institutional and governmental behaviors.

A detailed, extended case study of one unit is perhaps one of the most telling ways to identify changes that may occur over time as a result of rankings. Trow (1999) studied his own university's response to the first iteration of the NRC study wherein some of their biological sciences programs did not perform as well as expected. The results, spurred by steadfast university leadership and a commitment of faculty time and resources, led to substantial reorganization of the biological sciences on Berkeley's campus that were deemed successful after several decades had transpired. The convergence of external pressures and internal, institutional structures are a strong case example of how rankings studies, no matter how cautiously-accepted they may be, can lead to broad changes in institutional priorities and improvements in quality over time.

Harris and James (2006) provide an overview of two survey instruments used annually to assess graduates of all Australian universities. There is no systematic

research about the influences on institutional policies and practices related to the two survey tools, thus they provide their own observations. The data are especially useful because they highlight the value of teaching and learning outcomes and provide institutions with more information than they have ever previously had. In some ways this broader scope of assessment data can be likened to the NRC study as well. Survey data and results can be used by institutional management, students, policymakers, and the government to focus on information dissemination, quality assurance, and performance-based incentive funding. In Australia, the surveys appear to have begun influencing policy and practice within the higher education system now that a stronger market orientation and quantitative performance indicators have developed. Yet the impacts on quality assessment are difficult to prove with confidence, as is the case with the NRC study to date.

Similarly, Yorke and Longden (2005) interpret the performance measures used in the United Kingdom and maintain that the performance indicators can be used for impacting policy changes and decisions in higher education, including governmental policies over institutions. The report focuses mainly on explaining the indicators, correlations, data quality, and relative comparisons among participating universities but provides little empirical evidence for proving the impact.

Another report that showcases assessment and institutional reactions in Great Britain is the Higher Education Funding Council for England's 2008 study on league table structures. These tools are still uncertain mechanisms for concretely defining quality in higher education. Reputational factors still carry too much weight, and the compilation of the data and results could benefit from greater transparency.

Yet similar to the claims and hypotheses described at the NRC Convocation, this study also found that the practice of participating in league table projects has led to greater attention and data collection at higher education institutions. The league tables created impetus for making changes that might not have been made otherwise, but campus leaders also indicated such changes were already desired and the rankings themselves did not pressure these changes into occurring. Many reported implemented changes were in the areas of public relations, data collection and analysis, and the establishment of institutional key performance indicators. At the same time, the institutions do not feel they have sufficient input on the rankings and may in fact detract from other priorities and worthy policies, themes repeated in several of these in-depth studies.

The Institute for Higher Education Policy (2009) prepared an issue brief to summarize the rankings processes in existence and uses a case study format to describe the role rankings may play in institutional decisions. Rankings studies can influence many facets within an institution of higher education, such as strategic planning, organizational structures, and resource allocations. Positive results can be linked to the rankings studies, including improvements based on institutional data and comparisons and more attention to data collection and analysis. Institutional collaboration increases as a result of comparisons among peers. Yet negative outcomes may also result from rankings studies, such as when universities cater to the factors that inherently seem to lead to higher rankings, such as focusing exclusively on research, gaming statistics that have the potential to impact disadvantaged students, and making funding allocations that disproportionately

benefit the higher-ranked institutions. The key takeaways from this study are that rankings studies are used to differing levels and greater focus should be placed on the impacts of rankings among higher education institutions.

One of the few researchers to write extensively on graduate education and its funding, quality, and assessment is Ehrenberg. In one study, institutional responses to increased federal support for graduate students is reviewed to predict university behavioral responses (Ehrenberg, et al., 1993). They sought to determine if additional federal support for graduate education will simply induce universities to redirect its own resources differently across disciplines. This finding occurs only when external funding support changes are unexpected and transitory. Adjustments to changes in external support occur quickly at universities, especially at research extensive universities, when the external support changes are perceived as recurring or permanent and when the funding alters the distribution of funding type (e.g., RA vs. TA) to students. But the magnitude of overall change in terms of numbers of students supported is quite small. This work provides a model for looking at one discrete area of change based on comparative quality and assessment data.

One final, specific study found that university prestige – as measured by several of the rankings studies – does influence the number of annual licenses for university inventions (Sine, Shane, and Di Gregorio, 2003). Their work serves as another example of reviewing one key indicator and how quality studies may affect actions at a university. Next steps include broadening this review to determine how prestige and quality impact key decisions within a university or program and how to

make those decision-making practices informed by quality data and sustaining change over time.

Additional research highlights possible impacts of the rankings results, while at the same time acknowledges that minimal substantive changes occur (Miller, et al., 1996; Grunig, 1997; Sims and Syverson, 2000). Indirect effects are possible, such as the idea that rankings may influence peer-reviewed research proposals for funding, benchmarking for strategic planning purposes, strength of manuscript reviews for written publications, and increasing competition among peer institutions. Reputations for excellence in graduate education can lead to growth in research agendas for institutions, new faculty recruitments, and the enrollment of high-ability graduate students. Even with the various pitfalls for misuses of rankings well-documented, potential for positive effects on higher education quality do exist if the studies are well-designed (van Vught and Westerheijden, 2012). All these factors have the potential to create a cyclical effect that both purports to measure quality of a graduate program and enhances the quality of institutions, yet it is difficult to ascertain which must come first.

Evolutionary change theory findings and assumptions indicate that substantive change will not result merely from the existence of rankings studies. Instead, other institutional forces and peer effects must also be present to influence change and move university quality ahead. This dissertation aims to review this area, build on these theories, and address some of those influences through independent research.

Methodological Techniques and Considerations

Descriptive analyses, survey analysis, performance measurement, some regression techniques, time series studies, and case studies are all various methods utilized in higher education and assessment research. Using empirical data, surveys, and case study techniques, evaluations can review how institutions respond to rankings and quality studies and data. Various studies already discussed in this chapter can be synthesized via the chart below.

Table 2.1 – Synthesis of Methods Used for Evaluating Higher Education

Author(s)	Driving Questions	Methods
King, et al. (HEFCE)	Is the current rankings and assessment structure in the UK successful? How are results used?	Regression and factor analysis; online primary survey; case studies
Sine, Shane, Di Gregorio	How does university prestige influence technology licensing activities?	Generalized estimating equation regression
Feldman, Feller, Bercovitz and Burton	What factors influence the adoption of equity holdings as a technology transfer strategy?	Tobit regression
Geiger and Feller	What institutional factors influence the changes in share of in research funding?	Shift share analysis; descriptive analysis
Hazelkorn	How are higher education institution leaders reacting to league tables?	Surveys; case studies
Gioia and Thomas Dill and Soo Kivisto Meister-Scheytt and Scheytt Trow	-How do university perceptions of identity influence decision-making? -How do various rankings studies influence behavior and how can they be improved? -How do graduate schools manage university-government relationships in the context of agency theory? -What processes and individuals influence change management? -How does the pursuit of prestige	Case studies

	influence fiscal decision-making?	
Brewer, Gates and Goldman	-How do universities move incrementally toward core changes?	
Clark; IHEP	-What role do rankings studies have on institutional decisions?	

The Higher Education Funding Council for England's 2008 report on assessment practices, discussed in more detail above, combines surveying and quantitative results with case studies to highlight any impacts rankings efforts have on institutional decision making. Similarly, Hazelkorn's work over the past several years was a multi-phase survey and case study approach to study institutional decision-making using rankings. Her work provided a high level analysis of the opinions and uses of quality rankings among university leaders. Finally, the Institute for Higher Education Policy's 2009 issue brief and Trow's summary of change in the biological sciences at UC-Berkeley, among other studies, also use in-depth case studies to ascertain the role of rankings on institutional actions and decision-making.

Together, these research studies and reports provide a solid methodological foundation for further research in this arena. Primary data collection through surveys, including open-ended responses, provided the foundation for obtaining opinions on both the NRC study processes and its use on university campuses. Similar to these listed studies, the survey results led to the identification and cultivation for ideal case study sites to study these ideas in greater detail.

CHAPTER 3 – SURVEY METHODOLOGY

Institution-level change is difficult, especially sustainable and lasting change. The dissertation is focused on determining what influence national rankings studies at the graduate level have on institutional changes, including changes at the graduate program level. Both quantitative and qualitative elements are used to address two research questions: How do institutions respond to rankings and quality assessments? How does this potential for change affect the overall quality landscape for institutions of higher education?

To study these research questions, the dissertation uses survey design and analysis and qualitative case studies. Several research studies using these methods help inform and justify this choice (Hazelkorn, 2007, 2008 and 2009; King, et al., 2008; Sine, Shane, and Di Gregorio, 2003; Institute for Higher Education Policy, 2009) as appropriate for studying higher education use of assessment and rankings data. The National Research Council's Data-Based Assessment of Research Doctoral Programs (NRC study) methodology, participating institutions, dataset, and results are the foundation for this work.

Two similar versions of an original survey were run in the field from September through December 2011. The possible respondents were seven discrete populations, one at the central administration level for each institution that participated in the NRC study and six from select graduate disciplines. The

dissertation reviews the survey results data and discusses how decision-making and change in institutions of higher education may result from the participation in, and results of, national assessment studies such as the NRC. The primary goal of the surveys was to analyze how rankings studies have influenced organizational change by institutions of higher education, especially at the graduate education level. An overview of the survey design is the focus for this chapter. Chapter 4 then provides an overview and analysis of the survey results. The survey results were used to identify several case studies of higher education institutions and graduate programs to study in-depth, which is the focus of Chapter 5.

Survey Overview

The objective of the survey was to gather information from universities that participated in the NRC study. Each individual invited to complete the survey was from an institution of higher education that participated in the NRC study. Survey recipients include chief academic officers (i.e., the Provost or its equivalent) and faculty chairs/directors of graduate programs in select disciplines.

The University of North Carolina at Chapel Hill was excluded from the project to avoid possible conflicts of interest of a student studying her own institution. Also, due to this student's personal knowledge and participation in the NRC study response from UNC-Chapel Hill, it is inappropriate to include the campus in the data set for evaluation.

Two survey instruments, one for the central administration of each campus, and another for departmental/program chairs in select graduate disciplines that were included in the NRC study, were developed and implemented in summer and fall

2011. The surveys were administered using the Qualtrics web-based software through the Odum Institute for Social Science at UNC-Chapel Hill.

Institutional Review Board approval was requested and received for the survey portion of this project in the summer of 2011. The study number was 11-1442, which was initially authorized by the Behavioral IRB from August 9, 2011 through August 7, 2012 for the survey portion of the study. An extension was approved authorizing continued use of the data and results through July 22, 2013. It was determined that the risk involved to human subjects in this research was minimal.

Sample Selection Criteria – Central Administration

There are a total of 210 institutions in the NRC data set. Descriptive statistics for each of the 210 institutions is included in Table 3.1 on the following page. According to the current Carnegie Classification system ratings, there are a total of 297 doctoral serving institutions, including ratings at very high research activity (108), high research activity (99), and doctoral/research universities (90). The vast majority of doctoral serving institutions in the United States participated in the NRC study. All but one very high research activity university (Yeshiva University in New York) participated in the NRC study, and of the remaining institutions that did not participate, 24% were classified as high research and the remaining 75% were doctoral/research universities. There are 11 for-profit institutions included in the Carnegie ratings but not included in the NRC study.

All US institutions that are members of the Association of American Universities (AAU), generally considered to be the preeminent research institutions,

participated in the NRC study. A majority of US institutions that are members of the Association of Public and Land-grant Universities (APLU) also participated in the study. Emerging research campuses or some campuses with niche, professionally-oriented doctoral programs may have elected not to participate in the NRC study because their doctoral programs were not included in the taxonomy.

Table 3.1 – Institutional Overview for Surveyed Institutions

Characteristic	Total	Percentage
Public	141	67%
Private	69	33%
AAU – 59 total US	58	100%
APLU – 189 total	126	67%
Land Grant – 106 total	42	40%
US region – Northeast	54	26%
US region – Midwest	46	22%
US region – South Atlantic	37	18%
US region – South Central	35	17%
US region – West	38	18%

N = 210 institutions

The central administration survey was distributed to current chief academic officers from all 210 institutions. Contact and title information were collected from individual university websites during summer 2011. New and interim appointees were included as survey recipients. If the respondent asked for a designee to complete the survey on their university's behalf, the request was honored. Seven chief academic officers made such a request, most commonly asking if a knowledgeable graduate affairs dean or institutional research leader could complete the survey. The list of all survey recipients was loaded into the Qualtrics survey software for implementation.

Sample Selection Criteria – Graduate Program Level

The process used to identify survey recipients for the graduate program-level surveys followed a different logic. To study organizational change and quality improvements, the sample needed to include a broad range of individual programs representing physical sciences, biomedical sciences, social sciences, and arts/humanities. This dissertation project included one of each of the six NRC study broad fields[2] with graduate programs: Agricultural Sciences/Nutrition, Biological and Health Sciences/Neuroscience and Neurobiology, Engineering/Materials Science and Engineering, Humanities/English Language and Literature, Physical and Mathematical Sciences/Chemistry, and Social and Behavioral Sciences/Economics.

Six fields and 643 total programs are included in the data set. A total of 89% of universities (N=186) are represented with at least one of the selected graduate programs. The average number of graduate programs from the universe of universities is three, with a range from no programs up to 13 programs from one university. The number of survey recipients is not always equal to the number of programs due to joint programs, administrative structures such as co-directors or chairs overseeing more than one program, and programs that have consolidated since the NRC study occurred. Descriptive information for each of the fields and their survey recipients is included in Table 3.2 on the following page.

[2] The NRC study taxonomy had six *broad fields* identified. Within each broad field were a variety of 62 *fields* specific to each broad area. The third taxonomy layer was for each field to list individual *graduate program* names from the participating universities. Universities selected the NRC field within which their programs were most aligned for reporting and comparisons.

Table 3.2 – NRC Field and Recipients Overview for Surveyed Graduate Programs

Characteristic	Nutrition	Neurosciences	Materials Science	English	Chemistry	Economics
Number of programs	44	93	87	121	181	117
Number of recipients	44	98	82	122	174	116
Public	84%	62%	78%	68%	72%	65%
Private	16%	38%	22%	32%	28%	35%
AAU Percentage	50%	57%	49%	43%	38%	51%
APLU Percentage	86%	55%	80%	68%	72%	67%
Land Grant Percentage	59%	22%	36%	20%	27%	26%

N = 643 graduate programs
N = 636 survey recipients

Contact and title information for current faculty chairs/directors from each graduate program were collected from individual university websites during the summer and fall of 2011. New and interim appointees were included as survey recipients. If the respondent asked for a designee to complete the survey on their program's behalf, the request was honored. Only two program directors across all 643 programs made such a request. The list of all program survey recipients was loaded into the Qualtrics survey software for implementation.

Criteria for Selecting Fields

In evaluating which of the fields to select for study within the six NRC broad fields, a comparison was performed of those fields most highly represented across universities. The selection of fields and programs was designed to allow for the highest probability to survey as many institutions as possible. Table 3.3 shows the breakdown of the number of graduate programs in the selected six fields by all 210 universities in the dataset.

Table 3.3 – Graduate Programs in Six NRC Fields of Study by University

Number of Programs	Universities Represented	Percentage
0	24	11.4%
1	32	15.2%
2	33	15.7%
3	38	18.1%
4	31	14.8%
5	21	10%
6	23	11%
7	6	2.9%
8	0	0%
9	1	0.5%
10	0	0%
11	0	0%
12	0	0%
13	1	0.5%

N = 210 universities

Forty-three percent of universities have three, four, or five programs represented. The range increases to greater than six because universities could submit more than one program in a single NRC field. For example, one university may have two discrete PhD programs represented in the NRC Chemistry field, such as an Analytical Chemistry PhD and Organic Chemistry PhD in the same department or school.

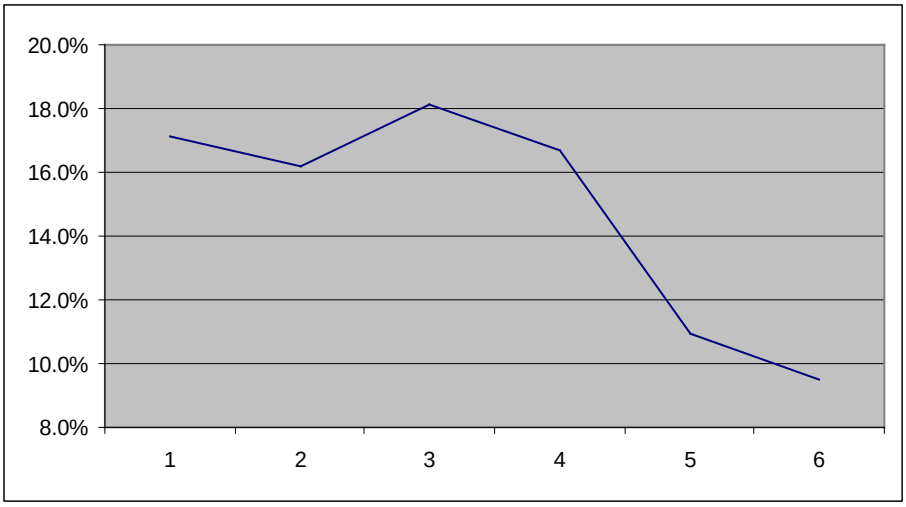
Also critical to evaluate was the number of universities represented across each of the six NRC fields, or how many opportunities does a university have to participate in the six program-level surveys. Approximately 46% of universities have programs included in three, four, or five fields. Twenty universities have programs included in all six fields. Table 3.4 shows the breakdown of universities with representation across the six NRC fields in the dataset. Below it is a graphical representation of the number of universities within each field.

Table 3.4 – Six NRC Fields of Study by University

Number of Fields	Universities Represented	Percentage
0	24	11.4%
1	36	17.1%
2	34	16.2%
3	38	18.1%
4	35	16.7%
5	23	11%
6	20	9.5%

N = 210 universities

Figure 3.1 – Six NRC Fields of Study by University



Universities Excluded from the Program Surveys due to Missing Fields

There are 24 universities that are not represented with graduate programs in any of the six NRC fields selected, thus they were excluded from the program-level survey phase of this project. This issue was unavoidable when selecting one NRC field within each broad field as no university had participating programs across every field. The universities that were excluded in the program surveys were still included in the central administration survey. These 24 universities had a total of 93 programs participate in some field of the NRC study, for an average of three programs per campus, but no programs in the selected study fields.

Care was taken to ensure the chosen fields allowed for a broad range of university participation. Nine of these 24 universities had only one total graduate program participating in the entire NRC study across all 62 fields. Two excluded universities only participated in the NRC study by virtue of joint programs with other participating universities, thus the data set includes at least the other partner from these joint programs. Several excluded universities participated in the NRC study by

virtue of a niche training mission for one or a few graduate programs, such as theological schools or medical schools.

The most common NRC program for which these excluded universities had representation in the NRC study is the general Biology category labeled to use only if the degree field is not specialized. Most other universities had Biology programs that were classified across other NRC fields.

There is one excluded university considered an extreme outlier because it had 15 programs participating in the NRC study across all fields, though none in the six fields selected. However, if the chosen NRC fields were switched to ensure the inclusion of this one university, additional universities in other fields would be excluded or not have the desired level of variety among fields.

Due to the reasons listed above, the dissertation project proceeded with the six selected NRC fields knowing these 24 universities would only receive the central administration survey without the opportunity to participate in any of the six program-level surveys.

Survey Design

There were two similar versions of the survey: one for the central administration at each institution that participated in the NRC study and one for faculty chairs/directors in the six doctoral program fields described above from the NRC study. The surveys were developed based on the literature reviewed earlier in this dissertation and with input and approval from dissertation committee members and experts from the Odum Institute for Social Science at UNC-Chapel Hill. Particular influence in question design and response choices came from literature on

research utilization (Weiss, 1979; Patton, 2000). Additionally, several of the usage examples and questions were derived from specific examples discussed at the 2011 NRC Convocation. The final versions of both surveys can be found in Appendix 3.1.

Topics in the surveys included questions about the NRC study, how the institution has used the results and data on their campuses thus far, how the institution used the data collection process for change in advance of the NRC study results being released, and what plans the institution has for further use of the results once they had been released publicly. Additional questions asked about the perceived usefulness for various aspects of the NRC study results, including the data collection process, outcomes data measures, and how these measures led to changes and improvements on their campuses and in their graduate programs. An open-ended question was included to obtain any additional feedback or comments on the NRC study and its perceived usefulness for enacting change. Finally, a question asked if the respondent would be willing to serve as a case study contact for additional interviews and questions.

The program-level surveys included some additional questions beyond those in the central administration survey. The graduate program directors were asked about their program's participation in, use of the results from, and perceived usefulness of the NRC study. They were asked to report on these topics for their institution as a whole as well. The ability to compare responses across levels was important for the data analysis phase, especially the ability to compare if the perceived usefulness by the chief academic officer mirrors that of their graduate program directors.

Survey questions were single-answer fixed response, multiple-answer fixed response, open-ended text, and Likert scale single-answer rankings questions. Respondents could skip any questions throughout the survey. Midway through the survey, respondents saw a statement that they were halfway complete to encourage continued participation. Only those individuals who responded they had used or have future plans to use the NRC study results on their campuses were shown questions about the specific areas for use. The respondents' names and email addresses were displayed within the survey to validate their accuracy, and if incorrect, the respondents were asked to provide correct contact information.

Survey Implementation

The survey was developed during the spring and summer of 2011. The release schedule occurred in waves during the fall 2011 semester using the Qualtrics software through the Odum Institute for Social Science at UNC-Chapel Hill.

A series of email communications were developed to invite survey recipients to complete the survey, beginning with a preliminary explanation of the survey. This preliminary communication allowed for validation of the proper contacts and introduced the survey idea in advance of requesting their time and participation. The survey invitation followed approximately one week later. If the recipient did not complete the survey within one week, a follow-up email communication requesting their participation was distributed. A final email reminder was sent a week later to any remaining non-respondents. Each email communication was sent via the Qualtrics survey software. The Qualtrics software tracks all recipients so that follow-up communications were only sent to those invitees who had not yet responded to

the survey. Each recipient received a personal email addressed to them, thus other invitees were not visible on the TO: email address line. Three communication attempts were expected to be sufficient to achieve a strong survey response and not upset possible respondents. The final versions of all communication templates can be found in Appendix 3.2.

It was anticipated that the central administration survey should take no longer than three minutes to complete. The program chair/director survey had additional questions, which lengthened the survey slightly, but it was still estimated to take approximately three to four minutes to complete on average. Estimated response times were validated through individual testing and feedback from survey experts at the Odum Institute for Social Science. These estimated times were folded into the communications to encourage busy individuals to spare a short amount of time to assist a student. The timestamps tracked within Qualtrics indicated wide ranges of times for survey activity, primarily because as soon as an individual clicked on their survey link, the active time begins. In general, no questions arose from respondents concerning the time spent on the survey.

The survey distribution and follow-up process took approximately one month to complete for each survey, from the initial email through the close date of the survey following the final reminder notice. The seven surveys were distributed in waves instead of all at one time to allow for focused attention and communications with each survey group. The times for each program survey communication were varied – one arriving in the morning, one midday, and one in the evening – to capture possible respondents with varying work habits at different times. Response

rate progression was tracked from one reminder to the next. The series of dates and times for each survey administration and communication point can be found in Appendix 3.3.

There were no identified direct benefits to participating subjects. As an incentive for completing the survey, respondents were offered a short report that provides an overview of the aggregated survey results from universities around the country. Responding universities were not identified in any way. The summary reports were distributed to all survey respondents in September and October 2012 as part of the case study identification process described further in Chapter 5. Copies of the summary reports can be provided by the author upon request.

Survey Response Rates

All seven surveys together initially achieved an average 48.4% response rate from 415 respondents (out of a total of 846) who completed and submitted the survey. The central administration survey was slightly above the overall average with a 49% response rate. For the program-level surveys, Materials Science had the lowest response rate at 39%, and three programs achieved a 52% response rate (Neuroscience and Neurobiology, Chemistry, and Economics).

The progression of reminder email communications boosted the overall response rates by almost half in most cases, with Materials Science experiencing the largest absolute increases from week to week.

Tables 3.5 and 3.6 on the following page show the submitted survey response rate progression in both percentages and raw number of respondents.

Table 3.5 – Submitted Survey Response Rate Progression (Percentages)

Survey	Email 2	Email 3	*Increase*	Email 4	*Increase*
Central Administration	27%	41%	*14%*	49%	*8%*
Nutrition	21%	41%	*20%*	48%	*7%*
Neuroscience and Neurobiology	36%	46%	*10%*	52%	*6%*
Materials Science	13%	28%	*15%*	39%	*11%*
English Language and Literature	23%	41%	*18%*	47%	*6%*
Chemistry	24%	44%	*20%*	52%	*8%*
Economics	28%	45%	*17%*	52%	*7%*
Averages	**24.6%**	**40.9%**	***16.3%***	**48.4%**	***7.6%***

Table 3.6 – Submitted Survey Response Rate Progression (Respondents)

Survey	Email 2	Email 3	*Increase*	Email 4	*Increase*
Central Administration	56	87	*31*	103	*16*
Nutrition	9	18	*9*	21	*3*
Neuroscience and Neurobiology	35	45	*10*	51	*6*
Materials Science	11	23	*12*	32	*9*
English Language and Literature	28	50	*22*	57	*7*
Chemistry	41	76	*35*	91	*15*
Economics	33	52	*19*	60	*8*
Totals	**213**	**351**	***138***	**415**	***64***
Averages	**30.4**	**50.1**	***19.7***	**59.3**	***9.1***

Table 3.7 below shows the final counts of survey respondents once valid partial responses that were captured in Qualtrics, but not officially submitted on the final survey page by the respondents, are included. Including all partial responses brings the total number of individuals who began the survey to 507. However, there were 53 fully blank responses – when the respondent opened the survey and began Qualtrics recording but did not complete any questions – that were removed from the final data set entirely. Once these blank responses were removed, the final total number of respondents in the study was 454 for a 54% response rate. The valid partial responses were included in the final analyses discussed further in Chapter 4.

Table 3.7 – Final Survey Responses with Valid Partial Responses Included

Survey	Final w/ partial responses	Response Rate
Central Administration	104	50%
Nutrition	23	52%
Neuroscience and Neurobiology	57	58%
Materials Science	34	42%
English Language and Literature	67	55%
Chemistry	99	57%
Economics	70	60%
Totals	454	53.7%

Survey Respondent and Population Comparisons

It is important to consider variation between the survey respondents and the total population of the survey sample, or non-response bias specifically. On the following pages, Table 3.8 provides an overview of the institutional respondents and program respondents from each of the six NRC fields. In Table 3.9 these data are compared against the total population samples found in Tables 3.1 and 3.2 above.

In general, the program survey respondents were on average just as representative as the total surveyed population of APLU members but less representative than the total surveyed population of AAU members. The central administration respondents reflected higher rates of both APLU and AAU membership than the total surveyed population. Public institutions also responded to the survey in slightly greater numbers than their private counterparts, especially for the central administration survey, relative to the total population of survey invitees for both groups.

The Neuroscience and Neurobiology program survey responses were the closest to their full program surveyed population with minimal apparent differences. Materials Science also differed minimally with the exception of its AAU representation. The percentage differences ranged somewhat higher for the Nutrition, English Language and Literature, Chemistry, and Economics programs and the central administration survey, each following the trends as described just above. None of the differences were higher than 9% for the central administration survey and 7% for any of the program surveys.

Given the low sample sizes for each survey, the differences can be attributed to relatively small numbers of respondents. Based on these reviews, it was determined that there is not bias in the survey respondents relative to the population of survey invitees.

Table 3.8 – Field and Respondent Overview for Survey Respondents

Characteristic	Central	Nutrition	Neuroscience	Materials Science	English	Chemistry	Economics
Number of responding programs	n/a	23	55	34	67	99	70
Number of respondents	104	23	57[1]	34	67	99	70
Public	75%	87%	61%	79%	63%	71%	70%
Private	25%	13%	39%	21%	37%	29%	30%
AAU Percentage	35%	48%	56%	44%	43%	35%	44%
APLU Percentage	69%	91%	53%	79%	64%	72%	70%
Land Grant Percentage	22%	65%	23%	35%	16%	32%	26%

[1] Two programs had both co-directors respond.

Table 3.9 – Difference between Respondents as Compared to Total Population of Possible Respondents from Tables 3.1 and 3.2

Characteristic	Central	Nutrition	Neuroscience	Materials Science	English	Chemistry	Economics
Number of possible programs	n/a	44	93	87	121	181	117
Number of responding programs	n/a	23	55	34	67	99	70
Number of possible respondents	210	44	98	82	122	174	116
Number of respondents	104	23	57	34	67	99	70
Public	8%	3%	-1%	1%	-5%	-1%	5%
Private	-8%	-3%	1%	-1%	5%	1%	-5%
AAU Percentage	7%	-2%	-1%	-5%	0%	-3%	-7%
APLU Percentage	9%	5%	-2%	-1%	-4%	0%	3%
Land Grant Percentage	2%	6%	1%	-1%	-4%	5%	0%

Limitations

Several limitations exist in the dissertation survey design. First, the survey data is obtained from only those institutions and programs that responded. Most notably there is a slightly lower response rate from private institutions and those who are members of the AAU, especially for the program-level survey, relative to the total population of surveyed campuses and programs. Although comparative analyses show minimal differences between the surveyed population and the respondents, these differences should be considered when reviewing the survey results.

As described in greater detail above, some universities were excluded from the program-level survey phase of this project because the institutions did not have participating graduate programs in the six chosen NRC fields. Thus the entire population of universities participating in the NRC study is not included in the program-level survey sample. This issue was mitigated to the extent possible in the careful selection of the six program fields. Additionally, every university that participated in the NRC study was invited to participate in the central administration survey.

Finally, it should be recognized that there may be a lack of knowledge or memory of the NRC study among survey respondents due to the length of time between the NRC study data collections in 2006 and its release and the dissertation project's surveys in 2011. This lag time is acknowledged and was addressed where possible in the specific case studies.

CHAPTER 4 – SURVEY RESULTS

The surveys conducted for this dissertation aim to provide results that will contribute to understanding how decision-making and change in institutions of higher education may be influenced by the participation in and results of national assessment studies such as the NRC. The literature reviewed in Chapter 2 provides evidence that the implementation of sustainable and meaningful change is difficult and requires certain themes to be present, such as key champions, institutional leadership, and accepted priorities toward continuous improvement. The survey results assist in showcasing campus reactions to the NRC study as one example. Analyses of the survey results show meaningful outcomes: while there is clear uncertainly about the validity of the NRC, many examples exist of universities and academic programs using the study data and results to assess their performance and improve themselves.

Results of the seven surveys discussed in Chapter 3 were analyzed with tools in the Qualtrics software. Data were also downloaded into CSV files for analysis in STATA and Microsoft Excel software. Presented below are results and discussion of the central administration survey, all six program level surveys combined, and several examples of individual program survey findings that were especially noteworthy. Detailed tables with descriptive statistics and frequencies from each of the surveys can be found in Appendices 4.1 through 4.9. A series of tables showing

the results from the central administration survey can be found in Appendix 4.1; the combined results from the six program-level surveys can be found in Appendix 4.2; the combined results from the high-quality program-level surveys can be found in Appendix 4.3; and the results from each of the disciplinary surveys can be found in the appendices following in order. The analysis will refer to these tables.

Central Administration Survey

The central administration survey was distributed in September 2011. There were a total of 104 useable surveys for a 50% response rate. On the question of whether the NRC study achieved its stated goals, 64% of respondents (n=67) reported it did not. At the same time, however, 52% of respondents (n=54) reported the NRC results have been incorporated, or there were plans to incorporate, in campus activities or discussions.

Responses from Table A4.1-1 show that the NRC results are used generally in areas of assessment and evaluation (e.g., general conversations about graduate education with n=39 and 75% of question respondents reporting this use, or program review with n=36 and 69% of question respondents reporting this use) more so than in areas of action or decision-making (e.g., budget or resource allocations with n=15 or 29% of question respondents reporting this use, or advocacy to state governments with n=5 and 10% of question respondents reporting this use). Program review and discussions about the quality of graduate education on campus are more evaluative in nature than advocacy activities, policy changes, or determining areas of growth for new graduate programs. These findings suggest the NRC results are used as a tool for considering the quality and future assessment of

graduate programs yet not relied upon heavily for decision-making on campuses. These findings are consistent with the survey open-ended comments regarding the use of the NRC study on campuses, as well as the findings discussed nationally since the results were released (NRC Convocation, 2011). This distinction from the survey was assessed further during the case study portion of this research.

The NRC Convocation on Analytic Uses in March 2011 provided evidence that universities found value in preparing, discussing, and collecting data for the NRC study. It was asserted that some universities found the preparation for the NRC study more valuable and meaningful than the results themselves. The central administration survey asked respondents about this idea, and the responses turned out to be very similar to the general use or plans for use of the NRC study on campuses.

By way of example, Table A4.1-2 shows the difference in responses between the question about use of the results generally and the question about use and value of the study in advance of the results being released. The results show fewer universities used the data for action or decision-making purposes prior to the study's release (e.g., six fewer respondents reported use of the NRC study for budget or resource allocations and five fewer respondents reported use for advocacy purposes in advance of the release as compared to the general use question). Five more campuses reported holding retreats or discussions to discuss important issues in graduate education, likely an indicator of preparatory activities for the NRC study release. But the findings do not show that many more campuses incorporated the preliminary data into decision-making activities or included them in evaluative or

assessment exercises such as program review. This finding suggests that those campuses with robust assessment and decision-making activity wanted to wait to see the actual study results before incorporating them into their campus processes. The NRC Convocation presentations and debate may have been reflective of those campuses who were growing their institutional research and data assessment capacities.

Data collection and analysis are the two areas where survey results showed major differences between responses about general/planned use currently and the question that asked about any uses in advance of the results being released. For these two choices, 16 more institutions reported engaging in data collection and data analysis efforts with the NRC study in advance of the results being released as compared to once the results were public. Faced with the somewhat daunting task of providing vast amounts of data to the NRC about their campus resources, faculty, and doctoral programs, universities likely realized gaps in their data availability and knowledge. Campuses took steps to collect and analyze more data about their people and activities, knowing the NRC study was pending. The implementation of these practices, assuming they were not one-time data collection mandates, suggests opportunity for sustainable change.

Universities could have used the NRC study definitions and data variables as a framework for the types of data they should collect on an annual basis, both to assess trends in these key national variables but also to prepare for the release of rankings on these variables. The release date of the NRC study results was pushed back several times. While exact numbers are not known, during this time, some

universities likely continued to collect data on their graduate programs. At the NRC Convocation, universities presented examples of new reporting efforts that developed on their campuses as a result of the NRC study. In one sense, preparation for the release could have become a continuous process on some campuses. In an effort to collect data to augment or refute potential NRC study results, universities had the opportunity to create institutionalized tools for data collection and assessment.

The remaining questions in the central administration survey asked respondents to rate their agreement with several statements concerning the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness. Tables A4.1-3 and A4.1-4 show these statements and the respondents' assessment of them.

Responses show that the processes involved in participating in the NRC study are rated more useful (n=59 for 60% of respondents) than the rankings that resulted (n=46 for 46% of respondents). The information included in the study's database/Excel spreadsheet was also found to be useful by 60% of respondents (n=61). Because there are few studies of this magnitude, preparing for the study, collecting the campus data, and subsequently having access to the data about all doctoral programs nationally appear to have contributed to the perceived usefulness of the NRC study. At the same time, the resulting rankings were not rated as highly useful as the database alone, a finding that introduces questions about the utility of this type of study, or at least the methodology and outcomes of the study.

A high number of respondents indicated their campuses had active participation among faculty and staff preparing for and participating in the NRC study (n=67 for 67% of respondents), including “champions” for the study (n=75 for 76% of respondents). This finding is in line with use theory that active involvement and champions for a project increase its perceived and actual usefulness. Because of the complexity and breadth of the NRC study, both in scope and in time, without campus leadership and engagement, its level of impact and usefulness would be significantly dampened.

A third of respondents (n=31) reported that the NRC study will improve the quality of graduate education on their campuses, an important finding. However, far fewer chief academic officers reported that the results persuaded them to implement change on their campuses (n=17 for 17% of respondents). Conversely, 40% of respondents (n=40) did agree with the statement that the NRC study and results have or would be used to justify decisions. The dichotomy between *persuasion* and *justification* to change was studied further in the case studies as described in Chapter 5.

Next, the dissertation turns to a review of the program-level surveys combined across all six fields presenting an overall picture and then an analysis based on the NRC quality rankings of programs. The NRC study was meant to assess program level quality, thus rolling up the data to create an institutional level rating of quality is not deemed wholly appropriate or supported by the NRC study methodology or data (NRC Report, 2010; NRC Convocation, 2011). Program

results were not aggregated in an attempt to identify top institutions or perform summary analyses at the university level.

Combined Program Surveys

The six program-level surveys were distributed in September, October, and November 2011. There were a total of 350 useable survey responses across six program surveys for an average 55% response rate. On the question of whether the NRC study achieved its stated goals, 68% of respondents (n=239) reported it did not. Yet 40% of respondents (n=141) reported the NRC results have been incorporated, or there were plans to incorporate them, in departmental and program activities or discussions.

Slightly different results are present in the program-level surveys from Table A4.2-1 when compared to the central administration survey detailed in Table A4.1-1. Programs used the NRC results for more tangible areas of action than the central administrators reported at the campus level. For example, budget requests were much more highly ranked as being influenced by the NRC study (n=62 for 45% of program survey respondents on this question) than budget allocations by central administration (n=15 for 29% of central administration survey respondents on this question). Graduate programs have also used the results for making decisions on admissions and recruitment activities (n=70 for 51% of question respondents), faculty hiring (n=37 for 27% of question respondents), and program policy revisions, such as funding decisions and mentoring policies (n=41 for 30% of question respondents).

Programs did report use of the results for evaluative activities, such as program review (n=91 for 66% of question respondents), campus discussions (n=53 for 39% of question respondents), and retreats to discuss important issues in graduate education (n=35 for 26% of question respondents), but not at the exclusion of some of the more action-oriented items on the list. Thus while use of the NRC results is more common for assessment and evaluation at a global campus level, doctoral programs themselves are just as likely to use the results in areas of action or decision-making within their programs. The divergence between the central administration and program survey responses was evaluated further in the case study portion of this research.

Table A4.2-2 shows the difference in responses between use of the results generally and the specific question about use and value of the study in advance of the results being released. There are several key findings to note in these results and when comparing the central administration survey and the aggregated program surveys. First, there are more instances of wide variation as compared to the central administration survey results. Many of the action-oriented uses of the NRC study, such as student admissions and recruitment activities (24 fewer programs reported use in advance of the results release) or faculty hiring and recruitment activities (23 fewer programs reported use in advance of the results release), could not occur until after the results were released. It appears programs used the results, likely the rankings more specifically, as part of these more tangible decisions or actions. Because some of the action-oriented uses are more relevant when making comparisons relative to peer or aspirant programs, these findings are expected.

Secondly, similar to the central administration survey, the discussions about assessing programs and the future of graduate education occurred at similar rates both in preparation for the NRC and once the results were released, with only two more programs reporting use of the results as part of campus discussions about graduate education in advance of the results being released. The evaluative outcomes seen from participation in the NRC study could occur regardless of timing and without necessarily having the results and rankings in hand.

Finally, when comparing these program-level responses to the central administration results in Table A4.1-2, there is less program emphasis on the data collection and analysis aspects that could occur prior to the results being released, with only two more programs reporting these uses, compared to the larger difference of 16 more campuses reporting these uses in advance of the results release in the central administration survey. It is likely that many of the data efforts on campuses are centralized in offices such as institutional research, which might only be reflected in the central administration survey results. However, it is still notable that the programs did not report similar focus and attention on how better to collect data and analyze themselves as a result of the NRC study participation. Program emphases appear to be more focused on peer comparisons and competitive uses.

The remaining questions in the program surveys asked respondents to rate their agreement with several statements concerning the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness. Tables A4.2-3, A4.2-4, and A4.2-5 show these statements and the respondents'

assessment of them. Several questions went further in depth than in the central administration survey, but comparisons are possible.

In general, programs found the NRC study results, including the various measures and components of the database (e.g., agreement to statements ranged from n=72 for 24% for the list of student campus resources to n=184 for 60% for the faculty productivity measures), more useful than the exercise of planning and participating in the study (n=146 for 47% of respondents). This finding is counter to what was found in the central administration survey where respondents expressed more support for the utility of the earlier phases of the study as compared to the results.

Programs found the measures surrounding diversity and student support more useful (n=178 for 58% of respondents) than the faculty counts (i.e., program size, which had n=135 for 45% of respondents) or student campus resources (n=72 for 24% of respondents). This finding is expected when taking into consideration how programs reported using the study results, such as comparing themselves against peer and competing programs on tangible measures like diversity counts and student funding. It appears the data were useful in large part because no true national database on these measures previously existed. Taken together with the more action-oriented usages reported above, such measures and data would be meaningful when programs intended to use the data for tangible outcomes such as student and faculty recruitment and resource requests.

When compared to the central administration survey responses, there is less agreement among program respondents for use factors such as having a champion

for the study at the program level (23% in program responses compared with 76% in the central administration responses). Programs still reported active involvement (n=126 for 42%) and an understanding of the NRC study (n=128 for 43%), but appear less likely than central administration respondents in investing time and resources in championing the study. This finding contributes to their perceptions of lower utility levels for the processes involved in participating in the NRC study.

Although a third of chief academic officers reported that the NRC study has or will improve the quality of graduate education on their campuses, only a quarter of program chairs agreed with this same statement about their doctoral programs (n=76). Again, this survey finding was explored further in the case studies. Similar to the central administration survey, there was a distinct difference between those programs that agreed the NRC study would *persuade* them to implement changes in their programs (only 15% agreed) compared to those who reported using the results to *justify* decisions (38% agreed), providing more indications that the study methodologies and areas for use need to be improved upon.

The final series of statements were meant to assess the programs' perceptions of usage and change broadly on their campuses as a result of the NRC study. Their responses can in turn be compared with the central administrator views on the same topics. The findings on the two questions about persuasion and justification of decisions were very similar to that of the central administration surveys with approximately the same percentages of respondents agreeing in both groups (around 16% and 42% agreeing respectively). However, more cynicism exists on the part of programs regarding whether the NRC study will improve the

quality of graduate education on their campuses. Approximately one-third of the chief academic officers responded that they agreed with this statement but only 21% of the program respondents (n=63) did so for their campuses. Although the programs tended to make more tangible decisions and take actions based on the NRC survey results within their programs, they did not have higher confidence that such actions would lead to quality improvement broadly on their campuses.

Survey Comparisons Based on Quality Rankings

The combined program survey responses were also reviewed in the context of the six fields' quality rankings from the NRC study. This portion of the analysis sought to understand whether use levels of the NRC study and results varied in programs deemed to be of high quality.

Assumptions about the reactions of such programs could be varied, especially because of the graduate community's response to the NRC study. On one hand, these programs may be more likely to value the study, even with its known flaws, because their own program fared well. If the study results validated their own successes through high-quality rankings, the programs may feel the study was well-designed and worthy of study and attention. Conversely, programs who fared well may be more apt to acknowledge their strong showing and promptly relegate the results to the back burner, deeming them unnecessary of additional attention because they were already so strong. Using Gormley and Weimer's theory of response to external pressures, it appears that appropriate, functional responses may be a combination of both acceptance of the study's methodology and direct use of the results to understand which aspects of their program led to the high-quality

results in an effort to enhance these characteristics to retain strengths. The results were reviewed to determine if these functional responses were found.

Additionally, a series of assumptions were required to identify which programs were designated high-quality for the purposes of this analysis, including which of the ranges of rankings to use as the foundation and which confidence level to use. The program-level quality indicator comes directly from the NRC study results. The 5th percentile for the program's overall rankings was used to look at the percentage of programs that are leaders nationally within each NRC field. Because of the higher acceptance levels in the graduate and higher education communities, based on media reports and the 2011 NRC Convocation discussions and presentations, as well as the more limited reliance on reputational factors, the NRC study's S range of rankings at the 5th percentile was used to designate programs as high-quality.

Each of the six fields was reviewed, and in line with the NRC study's methodology, programs were designated as those that could have fallen into the top 10% or the top 25% of all programs based on the 5th percentile in the S range of rankings. The number of programs deemed high-quality in each field varies based on the number of programs in a given field and the S ranges and also the mix of programs within a field. Some fields are much more heterogeneous than others. For example, the English Language and Literature field includes consistent degree programs in English, but the Materials Science and Engineering field is broader and has multiple variations of graduate programs included in it.

Out of the six fields and 643 total programs in the data set, based on this analysis, 114 programs (or 18% of the total) could have been in the top 10% of their

respective fields and 254 (or 40% of the total) could have been in the top 25% of their respective fields. These are the programs deemed high-quality for the purposes of this discussion. Of those high-quality programs, slightly more than half of them were also survey respondents, with 135 (or 21% of the total) responding to their program-level survey. Of the 135 high-quality program respondents, 57 programs (or 9% of the total and 42% of the high-quality subgroup) could have been in the top 10% in their respective fields. As would be expected, the 135 high-quality programs reflect higher rates of AAU membership with representation from 73% of the programs. Approximately 55% were from public institutions, 56% were members of APLU, and 24% were from land-grant institutions.

The survey results for programs in the top quartile of their field were compared against survey responses from the entire set of program responses. This comparison helped to determine whether perceived success in the NRC study methodology influenced the respondents' activities using the NRC study outcomes.

In comparing the 135 high-quality program survey responses to all combined survey responses, only slight differences are noted. On the question of whether the NRC study achieved its stated goals, 63% of respondents (n=85) reported it did not (as compared to 68% of all program respondents), indicating slightly higher levels of confidence in the study for this subgroup of high-quality programs. However, 48% of high-quality program respondents (n=65) reported the NRC results have been incorporated, or there were plans to incorporate them, in departmental and program activities or discussions. This reported rate of usage is higher than the 40%

reported for the combined program survey results as well as reported rates of use for most of the individual disciplines.

Table A4.3-1 shows the specific areas of use that respondents indicated the NRC study results have been, or will be, used in their doctoral programs. When compared to the combined program surveys detailed in Table A4.2-1, the high-quality programs had lower levels of use in many areas, especially the more assessment-oriented activities. Of particular note are lower reported responses on the focus on data collection (n=14 for 22% of question respondents compared to 26% of all program respondents on this question) and data analysis (n=9 for 14% of question respondents compared to 22% of all program respondents on this question) activities. Higher-quality programs, primarily based at AAU and research-intensive universities, may not have as strong a need to retool their data and assessment efforts on campus as a result of massive projects like the NRC. They may already have the infrastructure and central resources in place for data collection and analysis activities on campus.

High-quality programs did report slightly higher levels of use on program review (n=48 for 74% of question respondents compared to 66% of all program respondents on this question), budget requests (n=32 for 49% of question respondents compared to 45% of all program respondents on this question), faculty recruitment (n=17 for 26% of question respondents compared to 24% of all program respondents on this question), and student recruitment and admissions activities (n=36 for 55% of question respondents compared to 51% of all program respondents on this question). Most of these action-oriented activities could benefit

from positive results on rankings studies such as the NRC. An expected, functional outcome with positive results is to tout them when attempting to recruit students and faculty to graduate programs.

Table A4.3-2 shows the difference in responses between use of the results generally and the specific question about use and value of the study in advance of the results being released. With almost no differences between advance use and current plans for use, the high-quality program surveys report less emphasis on program review, accreditation, and the data collection and analysis focus areas than the aggregated program surveys show, which had more activity in advance of the results release. Many of the other trends are consistent with the combined program findings, which will be further detailed between the individual disciplines in the sections below.

The remaining questions in the program surveys asked respondents to rate their agreement with several statements concerning the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness. Tables A4.3-3, A4.3-4, and A4.3-5 show these statements and the respondents' assessment of them. In general, the high-quality programs were more favorable about the usefulness of the NRC study, both in the preparatory activities (n=62 for 52% as compared to 47% of all program respondents) as well as the various components of the study results. They report higher levels of active participation in the study (n=62 for 53% as compared to 42% of all program respondents), including higher rates of agreement with statements about having champions for the study (n=36 for 31% as compared to 23% of all program respondents). These findings are

in the line with functional response theory that those programs who performed well in the study would report higher levels of agreement with the study's methodology, strong engagement with the study, and their evaluation of the usefulness of the study results.

High-quality programs reported agreement at higher rates than the overall combined programs that the NRC study has or will improve the quality of graduate education in their own programs (n=35 for 30% as compared to 26% of all program respondents) and on their campuses broadly (n=30 for 26% as compared to 21% of all program respondents). Similar to the central administration survey and the overall combined program responses, there was a distinct difference between those programs that agreed the NRC study would *persuade* them to implement changes in their programs (only 14% agreed) compared to those who reported using the results to *justify* decisions (44% agreed), further evidence that methodological changes are necessary to lead to improved emphasis on continuous improvement.

The final series of statements were meant to assess the programs' perceptions of usage and change broadly on their campuses as a result of the NRC study. The high-quality program responses are slightly more favorable about these outcomes than the overall combined program findings. There were 18% of high-quality program respondents who agreed that their campus was persuaded to implement change as a result of the NRC study (n=21) as compared to 16% of all program respondents. Similarly, there were 46% of high-quality program respondents who agreed that the NRC results led to justification of campus decisions (n=53) as compared to 42% of all program respondents.

The hopeful, functional response was that high-quality programs may in fact be more apt to use the NRC study to improve themselves by studying their results and looking for ways to retain strengths in those characteristics that contributed to their high-quality showing. These outcomes were not fully shown in survey results. It appears the norm was usage levels just as consistent with that of all programs, slightly trending toward lower levels of use, yet with higher levels of appreciation for use of the study. Because they performed well, high-quality program may not have the impetus to rely on the study to compare themselves and look for opportunities to improve on multiple fronts. They also might not have as many perceived areas in which to improve, which would contribute to lower levels of use and a reliance on maintaining the status quo in their programs. Instead, the key areas of use appear to be more tangible activities that allowed for capitalizing on their successes, such as with faculty and student recruitment activities and budget requests to administrators. And due to these successes, the high-quality programs may in turn report greater levels of agreement with the usefulness of the NRC study.

Next, the dissertation turns to a discussion of the six specific program-level surveys and their results.

Individual Program Survey Comparisons

In this section, a brief overview of each program survey across the six NRC broad fields is presented. In discussing the survey results, primary focus is geared toward assessing similarities and differences between each discipline's responses as compared to the combined program trends and the central administration survey

findings. Many themes are similar across each of the disciplines; however, some important distinctions are also found.

A summary display of the data is presented in the following tables with specific descriptions to follow. Arrows and colors indicate the direction and magnitude of reported use, distinctions between current/planned use and use prior to the results release, and the level of agreement with the various themes about the NRC study. Upward arrows and greener colors indicate higher levels of use and agreement, while downward arrows and redder colors indicate lower levels of use and more disagreement with the statements. The colored grids provide an overview snapshot of comparisons across disciplines, which will be discussed in greater detail in the sections that follow. All detailed data tables for each program survey can be found in Appendices 4.4 through 4.9.

Table 4.1 – Summarized Areas of Use for Doctoral Program Survey Respondents

	Nutrition	Neuroscience	Materials Sci	English	Chemistry	Economics
Academic/curriculum revisions	27%	21%	36%	18%	27%	17%
Accreditation and/or assessment activities	18%	11%	18%	18%	16%	26%
Budget and resource requests to deans and/or administrators	27%	5%	45%	50%	56%	61%
Doctoral program policy revisions	45%	5%	27%	29%	36%	35%
Faculty hiring plans	18%	5%	55%	25%	36%	22%
Faculty recruitment	18%	11%	55%	14%	27%	30%
General conversations about key topics in graduate education within the campus	36%	58%	45%	36%	29%	43%
Identifying focus areas for future data analysis	27%	42%	36%	14%	16%	17%
Identifying focus areas for future data collection	36%	37%	36%	29%	22%	13%
Peer comparisons to identify your program's strengths and weaknesses	55%	42%	91%	46%	80%	48%
Program review	82%	58%	64%	46%	78%	70%
Specific retreats to discuss graduate education quality and/or future directions	27%	5%	73%	7%	33%	26%
Student recruitment and/or admissions	55%	47%	55%	54%	53%	43%

Table 4.2 – Summarized Difference in Areas of Use in Advance of Results Release for Doctoral Program Survey Respondents

	Nutrition	Neuroscience	Materials Sci	English	Chemistry	Economics
Academic/curriculum revisions	-1	0	5	1	-2	0
Accreditation and/or assessment activities	1	6	4	1	3	-1
Budget and resource requests to deans and/or administrators	-1	2	1	-9	-7	-3
Doctoral program policy revisions	0	6	2	0	-6	-1
Faculty hiring plans	-1	0	1	-4	-6	-2
Faculty recruitment	-1	-1	1	0	-6	-4
General conversations about key topics in graduate education within the campus	3	-1	1	4	0	-5
Identifying focus areas for future data analysis	0	-1	0	3	1	-2
Identifying focus areas for future data collection	0	4	0	2	-3	-2
Peer comparisons to identify your program's strengths and weaknesses	-2	4	2	0	-8	1
Program review	1	6	4	5	-34	-3
Specific retreats to discuss graduate education quality and/or future directions	-1	5	1	0	-10	-5
Student recruitment and/or admissions	-1	-4	2	-5	-12	-4

Table 4.3 – Summarized Doctoral Program Rankings of Usefulness of NRC Study Elements

Table 4.4 – Summarized Doctoral Program Rankings of NRC Study Use Factors within Program

Table 4.5 – Summarized Doctoral Program Rankings of NRC Study Use Factors within University

	Nutrition				Neuroscience				Materials Sci				English				Chemistry				Economics			
Table 4.3	SA	A	D	SD	SA	A	D	SD	SA	A	D	SD	SA	A	D	SD	SA	A	D	SD	SA	A	D	SD
We found collecting and submitting the campus data for the NRC in 2005-2006 useful.	5%	37%	53%	5%	0%	44%	39%	17%	6%	34%	41%	19%	7%	39%	40%	14%	1%	56%	34%	9%	5%	34%	51%	10%
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful.	0%	26%	68%	5%	2%	31%	52%	15%	6%	22%	47%	25%	3%	28%	50%	19%	1%	33%	55%	11%	3%	16%	71%	10%
We found the actual NRC database/spreadsheet useful.	0%	42%	47%	11%	0%	30%	46%	24%	9%	38%	38%	16%	9%	29%	36%	26%	5%	49%	31%	14%	7%	42%	42%	10%
We found the NRC illustrative ranges of rankings useful.	0%	42%	47%	11%	2%	31%	50%	17%	9%	33%	42%	15%	9%	30%	32%	30%	4%	40%	40%	16%	2%	45%	38%	15%
We found the faculty productivity measures (e.g., publications, citations, awards) useful.	16%	47%	32%	5%	6%	54%	22%	18%	21%	45%	24%	9%	9%	37%	46%	9%	15%	57%	22%	6%	10%	41%	43%	7%
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful.	16%	53%	26%	5%	10%	52%	28%	10%	6%	55%	27%	12%	7%	54%	28%	11%	7%	52%	31%	10%	3%	43%	46%	8%
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful.	11%	42%	42%	5%	8%	50%	30%	12%	3%	47%	41%	9%	9%	54%	30%	7%	3%	53%	34%	9%	10%	43%	44%	3%
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful.	5%	21%	63%	11%	2%	46%	32%	20%	3%	45%	39%	12%	4%	49%	40%	7%	3%	52%	38%	7%	5%	52%	39%	3%
We found the faculty counts and allocations useful.	0%	26%	63%	11%	0%	26%	54%	20%	6%	29%	52%	13%	5%	42%	42%	11%	8%	51%	32%	9%	3%	43%	44%	10%
We found the listings of 18 student activities and campus resources useful.	0%	21%	68%	11%	0%	28%	58%	14%	3%	31%	53%	13%	0%	23%	63%	14%	1%	23%	57%	18%	2%	13%	74%	11%
Table 4.4																								
My campus/program had active involvement among faculty and staff for the NRC study.	5%	16%	53%	26%	4%	38%	42%	16%	3%	41%	34%	22%	9%	49%	33%	9%	2%	42%	49%	7%	2%	28%	53%	17%
My campus/program had one or more champions for the NRC study.	5%	37%	32%	26%	4%	31%	39%	27%	0%	13%	59%	28%	7%	19%	52%	22%	1%	19%	65%	15%	3%	10%	64%	22%
My campus/program understood the NRC study methodology and results.	0%	53%	26%	21%	2%	30%	50%	18%	3%	31%	41%	25%	5%	42%	39%	14%	2%	39%	43%	16%	12%	37%	46%	5%
My campus/program was persuaded to implement change upon seeing the NRC study results.	0%	11%	68%	21%	0%	12%	54%	34%	0%	23%	43%	33%	0%	14%	60%	26%	1%	13%	72%	14%	2%	14%	64%	21%
My campus/program has used or will use the NRC study results to justify decisions.	0%	42%	47%	11%	0%	24%	44%	32%	0%	31%	47%	22%	7%	28%	39%	26%	1%	52%	33%	14%	5%	28%	56%	11%
The NRC study broadly has or will improve the quality of graduate education at my campus/program.	0%	37%	47%	16%	0%	14%	57%	29%	0%	29%	39%	32%	9%	18%	42%	32%	1%	23%	61%	15%	3%	26%	53%	17%
Table 4.5																								
My campus was persuaded to implement change upon seeing the NRC study results.	0%	16%	63%	21%	2%	16%	66%	16%	3%	6%	69%	22%	0%	18%	63%	19%	2%	14%	73%	10%	2%	12%	72%	14%
My campus has used or will use the NRC results to justify decisions.	0%	21%	58%	21%	4%	34%	48%	14%	3%	31%	47%	19%	0%	42%	37%	21%	5%	45%	40%	11%	4%	43%	44%	9%
The NRC study broadly has or will improve the quality of graduate education at my campus.	0%	32%	42%	26%	0%	8%	64%	28%	6%	19%	47%	28%	2%	23%	53%	23%	1%	21%	62%	15%	2%	19%	61%	18%

SA=Strongly Agree; A=Agree; D=Disagree; SD=Strongly Disagree

Nutrition

The Nutrition program survey was distributed in September and October, 2011. There were a total of 23 useable survey responses for a 52% response rate. On the question of whether the NRC study achieved its stated goals, 74% of respondents (n=17) reported it did not. However, 48% of respondents (n=11) reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the Nutrition survey results from Table A4.4-1 to the central administration survey from Table A4.1-1 and the aggregated program-level surveys from Table A4.2-1, some similarities are present yet so are some interesting differences. Similar to all other programs, Nutrition appears to have used the NRC results for more tangible areas of action than their chief academic officers reported at the campus level. For example, 55% programs reported using the NRC study through peer comparisons and 45% reported using them as part of policy revisions to their program. Similarly, Nutrition respondents reported higher rates of using the NRC study to recruit students (n=6 for 55% of question respondents) and enact policy changes in their doctoral programs (n=5 for 45% of question respondents) compared to the averages for all programs (51% and 30% respectively). Conversely, the field reported lower rates of using the NRC study for budget and resource requests (n=3 for 27% of question respondents compared with 45% for all program respondents on this question) and faculty hiring plans (n=2 for 18% of question respondents compared with 27% for all program respondents on this

question) or faculty recruitment activities (n=2 for 18% of question respondents compared with 24% for all program respondents on this question).

One specific area where Nutrition appears to be somewhat different than their fellow science-oriented disciplines is in their lower use of the NRC study for admissions and recruitment purposes (n=6 for 55% of question respondents), compared with 91% for Materials Science and 80% for Chemistry. One potential underlying reason for this difference may be that these Nutrition programs have greater access to national admissions data within public health fields[3]. Therefore, they may have less need for such data from a national survey such as the NRC study as compared to some of the other programs.

Table A4.4-2 shows the difference in responses between use of the results generally and the specific question about use and value of the study in advance of the results being released. The Nutrition findings are more in line with the central administration survey use responses than the combined program-level surveys. While the number of respondents for the Nutrition survey is small, there is not as much variation between the reported uses of the NRC study results and any changes that might have been implemented as a result of participating in the NRC study in advance of the results being released. There are higher reported rates of participating in campus discussions about graduate education (three programs reported this use in advance of the results release) and somewhat lower rates of

[3] Public health as an overarching discipline has a national admissions application called SOPHAS that is similar to other professional admissions portals. This application is a central clearinghouse for all applicants interested in any public health graduate degree areas (e.g., Nutrition, Health Policy, Epidemiology), although typically more so at the master's degree level. While not all universities participate in SOPHAS, having access to a centralized source of data about admissions may be part of the reason Nutrition programs do not need to rely on the NRC study data as heavily as some other disciplines.

peer comparison activity (two fewer programs reported this use in advance of the results release), which is similar to the central administration reported uses of the NRC study and results.

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.4-3, A4.4-4, and A4.4-5. The Nutrition program findings align with the combined program surveys responses, although their level of agreement is lower on most questions. Nutrition respondents found greater value in the NRC results for specific measures such as student support (n=10 for 53% of respondents), diversity matters (n=13 for 69% of respondents), and faculty productivity (n=12 for 63% of respondents). Again, the contrast with the central administration survey results is telling. The chief academic officers reported greater agreement with the usefulness of participating in the NRC study data collection efforts and preliminary activities on their campuses. Yet the programs reported greater usefulness from the actual results, although not necessarily the illustrative ranges of rankings.

Nutrition respondents reported much lower levels of active involvement in the NRC study in their departments (n=4 for 21% of respondents) as compared to the central administration survey (67%) or the combined programs responses (42%), contributing to their overall lower levels of use of the study. Fewer programs responded that they were persuaded to implement change upon seeing the NRC study results (n=2 for 11% of respondents), although a higher percentage of respondents agreed with the statement that their program will use the NRC results to justify decisions (n=8 for 42% of respondents). The 37% of respondents who

reported that the NRC study will improve the quality of their graduate program is actually higher than the central administration finding (32%) or the combined programs responses (26%), a somewhat contradictory finding given several of the Nutrition responses imply greater skepticism of the NRC study than their fellow disciplines.

Unlike their central administration and fellow discipline counterparts, Nutrition chairs did not report high rates of campus use to justify decisions (n=4 for 21% of respondents). However, the Nutrition survey did show a higher level of agreement on the NRC study's ability to improve the quality of graduate education at their campuses. One-third of chief academic officers and Nutrition chairs (n=6) reported agreement with this statement, although only one-fifth of all graduate programs agreed with this statement, indicating a belief in greater potential to enact change using the NRC study results, at least among the few survey respondents.

A key takeaway from the Nutrition survey is that its respondents reported less active involvement in and lower rates of agreement with the utility of various aspects of the NRC study, yet at the same time reported higher rates of agreement with the NRC study's ability to enact improvement and changes both in their doctoral programs and on their campuses.

Neuroscience and Neurobiology

The Neuroscience and Neurobiology program survey was distributed in September and October, 2011. There were a total of 57 useable survey responses for a 58% response rate. On the question of whether the NRC study achieved its stated goals, 81% of respondents (n=46) reported it did not. There were still 35% of

respondents (n=20) who reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the Neuroscience and Neurobiology survey results from Table A4.5-1 to the central administration survey from Table A4.1-1 and the aggregated program-level surveys from Table A4.2-1, both similarities and differences can be detailed. As a whole the Neuroscience and Neurobiology programs appear most cynical and least likely to use the NRC results when compared against their fellow disciplines. They reported the lowest rates of use on many program-specific factors, both on action-oriented options, such as program policy revisions (n=1 for 5% of question respondents), faculty hiring (n=1 for 5% of question respondents), and faculty recruitment (n=2 for 11% of question respondents), and on assessment and evaluative activities. Of particular note is the response on budget requests, which shows a large difference between the Neuroscience and Neurobiology respondents, reporting 5% usage, and their fellow disciplines, the lowest of which is 27% reported usage for Nutrition with an average of 48% for all five disciplines.

At the same time, the Neuroscience and Neurobiology respondents were more in line with the central administration survey respondents regarding the importance and use of the NRC study for the data collection and analysis focus areas. The programs reported higher rates of use in these areas at 37% and 42% respectively; but in sum did not report using the NRC study results widely on many tangible outcomes employing those data efforts.

In reviewing the reported use of the NRC study and data collection processes in advance of the release, as shown in Table A4.5-2, the responses show greater

variation than most other programs. Neuroscience and Neurobiology has the most positive variation, indicating greater rates of usage prior to the NRC study results being released (e.g., six more programs report use on program review, accreditation, and program policy revisions each in advance of the results release). Given their reported higher rates of use surrounding data collection and analysis efforts, this finding is expected as much of that activity could have occurred prior to the report release.

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.5-3, A4.5-4, and A4.5-5. The Neuroscience and Neurobiology program findings reflect similarities with the combined program responses. They reported lower rates of agreement in a few areas, but those were primarily on statements about the actual NRC study results as opposed to the processes employed on their campuses, which enjoyed greater rates of agreement (n=24 for 44% of respondents). These findings do not necessarily correspond to the strong cynicism described above regarding low usage on the specific areas, perhaps indicating recognized value for the NRC data and study results on an intrinsic level even if they are not incorporated into program decision-making in more tangible ways.

Where the program's skepticism does show clearly is in the levels of agreement about the outcomes of the NRC study. The Neuroscience and Neurobiology programs reported lower rates of agreement on both statements about persuading and justifying change at the program level (n=6 for 12% and n=12 for 24% of respondents respectively). Findings are similar at the campus level (n=9 for

18% and n=19 for 38% of respondents respectively). They also have some of the highest rates of disagreement with the statements that the NRC study will contribute to improvements in the quality of their graduate programs (n=7 for 14% of respondents) and graduate education broadly on their campuses (n=4 for 8% of respondents).

A key takeaway from the Neuroscience and Neurobiology survey is that while its respondents provided some positive agreement with the value of collecting data and analyzing themselves, in general there was little agreement with the value of the NRC study results or benefits to incorporating them in program or campus activities, discussions, and decision-making. The programs in this discipline, one of the most highly affected by the treatment of interdisciplinary programs in the NRC study, did not report utility or long-term possible program and campus improvements based on the NRC study or its results.

Materials Science

The Materials Science program survey was distributed in September and October, 2011. There were a total of 34 surveys useable survey responses for a 39% response rate. On the question of whether the NRC study achieved its stated goals, 71% of respondents (n=24) reported it did not. While it is the lowest among all survey disciplines, there were still 32% of respondents (n=11) who reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the Materials Science survey results from Table A4.6-1 to the central administration survey from Table A4.1-1 and the aggregated program-level

surveys from Table A4.2-1, some key differences should be noted. In contrast to the Neuroscience and Neurobiology program, Materials Science respondents are some of the most positive in terms of reporting use of the NRC study. They reported the highest rates of use on many program-specific areas, including retreats (n=8 for 73% of question respondents), peer comparisons (n=10 for 91% of question respondents), academic revisions (n=4 for 36% of question respondents), faculty hiring (n=6 for 55% of question respondents), faculty recruitment activities (n=6 for 55% of question respondents), and student admissions and recruitment activities (n=6 for 55% of question respondents). Aside from a few areas of assessment activity, these survey results indicate even higher levels of use than at the central administration levels.

In reviewing the reported use of the NRC study and data collection processes in advance of the release, as shown in Table A4.6-2, the responses show somewhat limited, though all positive, differences in usage rates before the results were released and those actions taken in preparation for the release. The Materials Sciences programs that did report greater usage of the study in advance of the results being released did so on the more action-oriented factors, such as making academic and curricular revisions to their programs (with five more programs reporting use in advance of the results release) and some activity around student and faculty recruitment.

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.6-3, A4.6-4, and A4.6-5. The Materials Science program findings indicate general

alignment with the overall combined program survey results with no outlying areas to note. Thus the heavily-reported use factors do not necessarily translate into increased agreement about the usefulness of various elements of the NRC study, at least not at higher levels than their fellow disciplines.

The one area of distinction in this program's survey responses relates to higher levels of agreement with the usefulness of the faculty productivity measures (n=22 for 67% of respondents with 21% strongly agreeing, as compared to 60% of all program respondents with only 12% on average strongly agreeing). Because Materials Science is such an interdisciplinary field, it is possible that comparative data on faculty scholarship is not as widely accessible as it might be in other, more traditional disciplines. Thus, the NRC study did appear to provide meaningful data for this particular area, which then translated into more action-oriented uses at the program level.

A key takeaway from the Materials Science survey is that its respondents reported much higher levels of use of the NRC study and its results on most use factors than virtually all other disciplines and the central administration. That result occurred even though this program was the least likely to report incorporating the NRC study into their programs' planning and decision-making. It appears those that did incorporate the results did so purposefully. That said, their use levels did not necessarily translate into increased agreement on the usefulness of the aspects of the NRC study or processes on campus, except for the area of faculty productivity data.

English Language and Literature

The English Language and Literature program survey was distributed in September and October, 2011. There were a total of 67 useable survey responses for a 55% response rate. On the question of whether the NRC study achieved its stated goals, 69% of respondents (n=46) reported it did not. At the same time, however, 45% of respondents (n=30) reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the English Language and Literature survey results from Table A4.7-1 to the central administration survey from Table A4.1-1 and the aggregated program-level surveys from Table A4.2-1, the results show this program is not truly an outlier in any particular area. Rather, they appear to be in line with their peer disciplines as users of the NRC study and its results. As is the case at the aggregated program level, they report more use on the action-oriented factors, such as student recruitment (n=15 for 54% of question respondents) and budget requests to administrators (n=14 for 50% of question respondents). Counter to the central administration respondents, the English Language and Literature responses indicate less emphasis on assessment activities like program review (n=13 for 46% of question respondents), peer comparisons (n=13 for 46% of question respondents), and data collection (n=8 for 29% of question respondents) and analysis efforts (n=4 for 14% of question respondents).

In reviewing the reported use of the NRC study and data collection processes in advance of the release, as shown in Table A4.7-2, the responses show some

variation between use of the results once they were released and what could occur prior to the release but not large differences (e.g., five more programs reported use of the results with program reviews, but five fewer programs also reported use of the results for student recruitment activities). Similar to the Nutrition program results, because most of the reported areas of usage in English Language and Literature are more action-oriented, these programs needed to have the results in hand before taking action or finding utility in the results, such as with budget requests and recruitment activities.

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.7-3, A4.7-4, and A4.7-5. The English Language and Literature program findings are mostly in line with the combined program findings with the exception of higher reported involvement in the NRC study (n=33 for 58% of respondents as compared to 42% for all program respondents). The responses also reflect slightly higher rates of agreement with the statements concerning improving the quality of their own graduate programs (n=15 for 27% of respondents) and graduate education more generally on their campus (n=14 for 25% of question respondents) as compared to all program respondents (26% and 21% respectively).

A key takeaway from the English Language and Literature survey is that its respondents reported higher rates of active involvement in their programs in the NRC study processes, which may have contributed to their slightly higher ratings on the NRC's potential to increase the quality of their graduate programs. However, no

key area of use stood out for this particular program as compared to their fellow disciplines.

Chemistry

The Chemistry program survey was distributed in October and November, 2011. There were a total of 99 useable survey responses for a 55% response rate. On the question of whether the NRC study achieved its stated goals, 67% of respondents (n=66) reported it did not. At the same time, however, 46% of respondents (n=45) reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the Chemistry survey results from Table A4.8-1 to the central administration survey from Table A4.1-1 and the aggregated program-level surveys from Table A4.2-1, some similarities and differences are present. Similar to the combined program survey results, the Chemistry program findings reflect more use on action-oriented factors such as peer comparisons (n=36 for 80% of question respondents), budget requests (n=25 for 56% of question respondents), and policy revisions within their doctoral programs (n=16 for 36% of question respondents). They did report lower levels of use on assessment activities such as discussions about graduate education (n=13 for 29% of question respondents) and data collection (n=10 for 22% of question respondents) and analysis efforts (n=7 for 16% of question respondents), which again is different than the central administration respondents' emphasis on the use in assessment-oriented areas.

In reviewing the reported use of the NRC study and data collection processes in advance of the release, as shown in Table A4.8-2, the responses show some

large discrepancies as compared to other programs. Chemistry reported much lower rates of use on a variety of factors prior to the results being released when compared against all program respondents. Program review, retreats, student recruitment activities, faculty recruitment and hiring planning, and doctoral program policy revisions all had much lower reported levels of use prior to the results release than other program survey respondents. Program review, in particular, saw 34 fewer programs report early use in this area as compared to current/planned use rates. Some, though not all, of these findings can be attributed to the fact that the results needed to be in hand before taking action. Because Chemistry reported such high rates of use on peer comparisons, the fact that these activities could not occur prior to the results being released is logical.

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.8-3, A4.8-4, and A4.8-5. The Chemistry program findings show higher levels of agreement with a majority of the use statements, including the various elements of the NRC study results, preparatory activities and engagement (n=52 for 57% of respondents), and the justification of changes in their programs (n=46 for 53% of respondents) and on their campuses (n=42 for 50% of respondents) as a result of the NRC study as compared to their peer disciplines. Like other programs, they were consistent in having only about a quarter of respondents report levels of agreement with the NRC study improving the quality of graduate education in their programs (n=21) and on their campuses (n=19), again suggesting room for

improvement in how similar data and study results could be tailored to improve continuous improvement efforts.

A key takeaway from the Chemistry survey is that its respondents reported relatively high rates of using the NRC study results on action-oriented use factors, such as peer comparisons to look for areas of strength and improvement. They also tended to agree more often that the various aspects of the NRC study results and the processes on their campus had utility. However, unlike their fellow programs, they did not report as much use prior to the release of the study results nor did their perspectives on the utility of the elements of the study lead to higher rates of agreement on the potential for improving the quality of their graduate programs or graduate education on their campus more broadly. These findings suggest the will to use the results for program improvement purposes is there, but the NRC study itself did not prove the best tool for doing so.

Economics

The Economics program survey was distributed in October and November, 2011. There were a total of 70 useable survey responses for a 60% response rate. On the question of whether the NRC study achieved its stated goals, 57% of respondents (n=40) reported it did not. There were still 34% of respondents (n=24) who reported the NRC results have been incorporated, or there were plans to, in departmental and program activities or discussions.

In comparing the Economics survey results from Table A4.9-1 to the central administration survey from Table A4.1-1 and the aggregated program-level surveys from Table A4.2-1, both similarities and differences can be detailed. As one of the

lowest overall reported users of the NRC study results, Economics is somewhat mixed in their survey feedback. For example, they reported comparatively higher rates of usage on budget requests to deans and administrators (n=14 for 61% of question respondents) and accreditation activity (n=6 for 26% of question respondents) but comparatively lower rates use on other action-oriented factors such as peer comparisons (n=11 for 48% of question respondents), student recruitment activities (n=10 for 44% of question respondents), and data collection (n=3 for 13% of question respondents) and analysis activities (n=4 for 17% of question respondents). There is a wide range of use in these survey respondents, and generally reported levels of use on individual factors were above that of their combined fellow disciplines.

In reviewing the reported use of the NRC study and data collection processes in advance of the release, as shown in Table A4.9-2, the responses show minimal variation. However, the variation that does exist is negative in that programs who reported use of the NRC study generally reported lower usage rates in advance of the results being released, though not to the extremes as seen with Chemistry. Of note, Economics reported lower use on a few key areas where central administrators indicated greater advance use, such as general discussions about graduate education (with five fewer programs reporting use in advance of the results release) and data collection and analysis activities (with four fewer programs reporting early use in these areas).

The statements about the usefulness of various aspects of the NRC study as well as factors contributing to its perceived usefulness are found in Tables A4.9-3,

A4.9-4, and A4.9-5. The Economics program findings suggest lower levels of active involvement among the programs in the NRC study (n=17 for 30% of respondents) as compared to 42% among all programs, but they also have general alignment with the combined program responses concerning areas of perceived utility. They were especially in agreement about the usefulness of the faculty productivity measures (n=31 for 51% of respondents) and student recruitment and admissions factors (n=35 for 58% of respondents), both of which would be key inputs to the factors reported with the highest use by the discipline. Similar to many of the other disciplines and the central administration respondents, there is greater agreement with the statements about the NRC study's ability to justify changes (n=25 for 47% of respondents) more than to persuade change (n=8 for 14% of respondents) or lead to quality improvements in graduate education on campus (n=12 for 21% of question respondents).

A key takeaway from the Economics survey is that its respondents reported lower rates of use of the NRC study, though in areas where they did use it they saw value, particularly in the faculty productivity measures. As a whole, the discipline did not support the NRC study's participation and results or see value in using the study data or outcomes. But again, this discipline's survey responses indicate some intentional pockets of use, which with the proper efforts, did find value in the outcomes.

Next the Chapter turns to an overview of the open-ended responses from several questions in the surveys, including an overview of the questions and categorizing the types of responses seen at both levels.

Open-Ended Responses

There was one open-ended survey question asking respondents for any additional thoughts on the NRC study and its use on their campuses. This question was included in both the central administration survey and the program-level surveys. There was a second open-ended question in the program-level surveys that asked respondents to provide any additional context to explain their level of agreement with the various statements about usefulness.

There were a total of 143 comments gathered across all of the program surveys and 36 comments gathered from the central administration survey. The comments were reviewed and categorized into several key themes as found in Table 4.6. The percentages reflect the rate each theme was present in the open-ended responses, not the total number of survey respondents. Open-ended comments will not be attributed to individual respondents, universities, or graduate programs in the dissertation discussion. General trends and observations from the open-ended responses are described further below.

Table 4.6 – Reported Open-Ended Survey Response Themes

Response Theme	Central	Programs
Assessment activity occurs already regardless of NRC	14%	9%
Contributed to another study or data source	8%	2%
Findings did not align with expected outcomes	0%	6%
Issues with data – collection and reported	19%	20%
Issues with the release date and study timing	58%	19%
Overall positive sentiments about the study	25%	8%
Reported lack of awareness for study	8%	13%
Study methodology was too complicated	45%	28%
Study was not useful to their specific program	n/a	7%
Will lead to caution about future/similar studies	6%	4%

Respondents were generally respectful of the National Academies and the NRC study, though largely due to appreciating the effort that went into the study more than the outcomes of the study. The results and the final study outcomes were generally not well-received in individual comments. Even those comments that were positive in nature – 25% of central administration comments and 8% of program-level comments – usually had other points in the same remark that were more constructive in nature. For example:

> "Unlike others here, I liked the R and S approaches, including confidence intervals. Some others argued about which weighting factors used to provide the final score, but I agree that some entity must create that and it was done a priori so was not biasing. In contrast, some of the survey data were too rigid. ... I also think that the NRC approach hurts interdisciplinary programs because it forces each faculty member to total 1.0 FTE. This makes sense for a survey, but it is negative for interdisciplinary programs."

> "While the data used were woefully out of date by the time of release and the methodology confusing and the revisions further undermined credibility, my sense is that the rankings (in both ranges) have and will be used to identify weak programs, encourage strong programs, and to allocate resources dedicated to graduate education. This is not because the NRC study was precise--it isn't--or up-to-date--far from it--but only because it can provide a very rough gauge of quality and productivity in the absence of other, better studies."

Comments centered on the lag time that occurred between data collection and the study release. This complaint was the most common theme among central administration respondents, mentioned in 58% of the comments, and the second most common theme in the program-level comments, appearing in 19% of the comments. The NRC authors acknowledged this lag time but argued it was not meaningful in a real sense because of how slowly graduate programs change (NRC Report, 2010; NRC Convocation, 2011; Glenn, June 2010). They believed that in general graduate programs and their faculty are relatively stable, especially at top

quality institutions. But this argument did not coincide with the comments in the survey responses, which in some cases were quite harsh in this regard. Whether or not it should have, the lag time significantly impacted the higher education community's level of comfort with the validity of the NRC study outcomes.

The study methodology and illustrative ranges of rankings themselves are not generally perceived to be useful or meaningful. Of note, several respondents commented on how the very complexity that made the NRC study a potential national gold standard in assessing graduate education led to its difficulties in understanding and awareness among its consumers. An "impossibly complex approach" was one respondent's opinion regarding the lack of usefulness of the NRC study, based in large part on the methodology's inability to be easily explained and used. Central administration respondents referenced the study's complexity in 45% of the comments; it was the most referenced theme in 28% of program-level comments. Respondents appeared to yearn for a simpler, easily understood methodology and outcomes – specifically wanting an ordinal ranking on which to base decisions and publicize their programs as opposed to the ranges of rankings on various measures that were deemed more-easily manipulated.

With the caveat of the lag time issues, the raw data were generally found to be useful, though not necessarily attributed to the NRC's success or with full confidence in their accuracy. Flaws are present in the NRC study data, particularly for interdisciplinary programs. But the scope of the data available to universities and graduate programs is unlike other databases or sources widely available nationally. Many reported that the NRC data was just one study or database among many that

they use for self-assessment and comparative purposes. Assessment is ongoing and clearly taken seriously at campuses and within graduate programs. One respondent acknowledged the flaws, yet said the NRC study shows it is possible, and responsible, to make management decisions based on real data. This type of response is an important show of support for perceiving the NRC study as an information instrument for implementing change and improvement in higher education.

Respondents commented on how the data collected are now used on their campuses. Some institutions reported not previously tracking some of the quantitative metrics on their graduate programs before the NRC study. Now they do so, including a few responses that indicated their campuses undertook efforts to perform their own internal reviews to augment or supersede the NRC study. Several respondents commented on their program's ability to show changes and a positive trajectory in quality measures, even collecting their own data post-NRC submission to show how their programs had improved in the intervening five years. There was a sense among respondents that having actual data on the quality of doctoral programs meant long-standing reputations – positive or negative – were not the only factors available for consumption in the academic marketplace. One comment summed this point well:

> "This most recent NRC study with its more objective measures of quality significantly bolstered the morale within my department in particular and our university in general. I think that for too long measures of quality in graduate programs were based on reputations established years ago and not necessarily maintained. That situation meant that very strong, up and coming programs could not make inroads into the established patterns. The new NRC rankings based on a variety of factors allowed very strong, but under respected programs to move into more appropriate positions in the rankings."

Several comments indicated some of the data collection and analysis has been outsourced to third party providers of university data intelligence. One respondent went so far as to state an opinion that the rise in these companies, specifically Academic Analytics, makes it unlikely there is a need for future NRC studies. This finding aligns with the comments from respondents who were concerned with the NRC study because it allowed for such wide variation in variable definitions and faculty allocations. For example, different universities could report on the same variable using different definitions, which thus impacted the outcomes. With a third party provider or another entity coordinating the data collection and analysis efforts, such issues become less problematic because every institution should be using the same definition for participating faculty and specific variables.

Another thread from the comments is that the discussions that were generated as a result of the NRC study were deemed useful. Even in the most extreme view – the study methodology was flawed, the data had inaccuracies, and the rankings were impossible to understand – the project generated discussions that were deemed valuable in graduate programs and across campuses. The resource trade-offs in time and attention for what is generally perceived as a late and failed study created much of the animosity toward the NRC study. But shining a light on the quality of graduate education, and the metrics that attempt to measure that quality, proved valuable for those respondents whose programs and campuses devoted time and attention to them.

While impossible to measure quantitatively, an interesting question would be to ask hypothetical opinions on how the NRC study *might have been used if* there

was greater transparency and timeliness on several factors that make the current iteration unacceptable, such as improvements in methodological decisions and data on faculty counts and publication activity, along with the obvious faster release of results. Several comments overtly stated or implied it was “a shame” that the NRC study was released so late and was so complex. These comments indicate there is general agreement behind the idea and spirit of the NRC study. There is perceived value in using data for decision-making and future quality improvements. One specific quote stated:

> “The delay in releasing data made the study fairly unhelpful, although of course we were happy to tout our programs that fared extremely well. The study was also useful as a trigger to emphasize graduate education and its importance on our campus, and this, I believe, will be the lasting benefit rather than any specific numbers from the ranges of rankings.”

This comment solidifies the general perspective on which the surveys, case studies, and this dissertation was based.

CHAPTER 5 – UNIVERSITY CASE STUDIES

This dissertation reviews decision-making and change implementation at campuses as a result of participating in national assessment studies such as the NRC. The previous two chapters described the results of surveys distributed to chief academic officers and program chairs and directors in select graduate disciplines at institutions of higher education that participated in the NRC study. Given the in-depth nature required to study institutional change and possible outcomes, case studies were chosen to allow for gathering more insights and information than the survey responses alone provide (Gioia and Thomas, 1996; Dill and Soo, 2005; Kivisto, 2007; Holmes, 2010; Meister-Scheyett and Scheyett, 2005; Brewer, Gates and Goldman, 2002; Clark, 2003; Trow, 1999). Additionally, several model studies effectively employ a combination of surveying and case studies to describe institutional change, some with particular emphasis on the influence of quality and prestige (King, et al. (HEFCE), 2008; Institute for Higher Education Policy, 2009; Sine, Shane, and Di Gregorio, 2003; Hazelkorn, 2007, 2008, 2009, and 2011).

The survey portion of this project was covered by the UNC-Chapel Hill Institutional Review Board approval under study number 11-1442. Minor changes related to the case studies and a project renewal were both approved by the Behavioral IRB in July through September 2012, with a revised expiration date of July 22, 2013. A consent release form was developed for the case study

interviewee subjects, which included agreement to be digitally-recorded during the interview; all but one interviewee agreed to be recorded. It was still determined that the risk involved to human subjects in this research was minimal.

Case Study Survey Responses

The surveys described in the prior chapters asked respondents if they would be willing to expand their responses further via participation in a confidential case study. Table 5.1 below shows the number of respondents from each survey who agreed to be contacted for additional information as part of a case study, including the percentage of all respondents this number represents. No single university responded to each survey and agreed to serve as a case study in all responses.

Table 5.1 – Respondents Agreeing to Case Study

Survey	Number	Percentage
Central Administration	22	21%
Nutrition	5	22%
Neurosciences and Neurobiology	14	25%
Materials Science	6	18%
English	7	11%
Chemistry	23	23%
Economics	8	11%

It was not realistic that each of these willing respondents could serve as a case study, both due to time and resource constraints as well as the dual-level nature of the surveys. For example, some university chief academic officers did not agree to serve as a case study even though some of their programs did agree, and vice versa. It was determined that the wishes of the chief academic officers at these institutions would be respected and taken as first priority. In instances where each of the program respondents did not agree to serve as a case, even though their chief

academic officer did agree, negotiation of the case study experience occurred carefully and transparently.

As described further below, three universities were selected to study further in terms of their use of the NRC study and results as well as their broader use of assessment data on their campuses. The unit of analysis was an important factor in the decision of cases. Given that this dissertation reviews both institutional level and program level change, and the NRC study results are more program-specific, care was exercised in making assumptions about the institution as a whole. To ensure that this dissertation reflected as broad a perspective as possible, interviews occurred with other central administration individuals at the case study institutions beyond the chief academic officer respondent. In addition to the chief academic officers and select graduate program chairs/directors, individuals in graduate colleges, dean's offices, or institutional research offices were interviewed.

Because case studies by nature are not representative of an entire population, it was determined that having a variety of university characteristics was more important than identifying required criteria for each selected site. For example, university status and groupings, such as public/private and AAU or land grant status, were varied among the final three cases. Other factors such as geographical dispersion and size were also varied, as opposed to selecting only large institutions or those similarly-situated near one another.

Case Selection Overview

The goal of the case studies was to study planning, data collection efforts, campus discussions, and processes in greater detail to determine how change on

campuses could result from rankings and assessment studies, such as the NRC, and to further understand variations in use within the central administration and program levels. Findings from the chosen case study universities were augmented through the use of news releases and public websites where appropriate as it relates to assessment and change.

The selection of case study universities occurred via two distinct but related analyses. The first analysis reviewed all universities based on responses to two questions from the surveys. Both the central administration survey and the program surveys began by asking respondents if they believed the NRC study met its primary stated goal[4]. Admittedly, the respondents' level of agreement with the stated goal can be influenced by a myriad of factors, not the least of which is the perceived usefulness and success of the outcomes. Next, each survey asked respondents if they have incorporated, or have plans to incorporate, the results of the NRC study in their campus or program activities and discussions. Responses to both of these questions can lead to a natural breakdown or grouping among respondents. The matrix in Table 5.2 shows the target responses for identifying three universities as possible case studies using the central administration survey responses.

[4] The specific goal that was presented to respondents came directly from the NRC study website and report. The exact goal statement presented in the survey was: "to provide an unparalleled dataset that can be used to assess the quality and effectiveness of doctoral programs based on measures important to faculty, students, administrators, funders, and other stakeholders".

Table 5.2 – Matrix of Case Study Possibilities Based on Use and Goal Responses among Central Administration Respondents

	NRC Study Met Goal	NRC Study Did Not Meet Goal	*Total*
Use of NRC Study	Case Study #1 (24)	Case Study #2 (27)	*54*
No Use of NRC Study	Case Study #3 (11)	No Case Study (40)	*52*
Total	*35*	*67*	

Note: totals do not sum equally due to response rates on individual use questions.

Case study #1 would be a university with a central administrator who believes the NRC study met its stated goals and reported using the NRC study outcomes. Case study #2 would be a university with a central administrator who did not believe the NRC study met its stated goals yet also reported using the NRC study. This outcome could lead observers to wonder why the results have been used on the campuses when in fact the intent of the study was not met. Case study #3 would be a university with a central administrator who indicated the NRC study met its stated goals but has not used, and does not report plans to use, the NRC study outcomes. This contradiction in responses was discussed further in the interviews.

Universities Excluded from the Case Studies

The fourth quadrant in the matrix is comprised of universities whose central administration respondents indicated the NRC did not meet its stated goals and reported no plans to use the NRC study outcomes. This finding appears logical on the surface. While there could be various reasons for why institutions did not believe the NRC study met its goals, or why they have chosen not to consider implementing the study results for any purposes on campus, it was determined that these

institutions did not merit additional in-depth study as part of this dissertation research.

This group of universities to be excluded from case studies was reviewed to confirm that other reasons that might have inhibited their use of the NRC study, such as low quality outcomes, were not readily present. Given the research profile of many institutions falling into this category – nine AAU institutions, 11 land grant campuses – low quality status, lack of graduate program depth and breadth, and low numbers of engaged faculty researchers do not appear to be reasons for not using the NRC study. Given that this dissertation is intent on better understanding what factors contribute to effective use of rankings studies and results, further evaluation of ones not using the study is not warranted via the in-depth case studies. Additionally, all but seven of the central administration respondents falling into this fourth quadrant did not agree to serve as case studies. The survey results alone, particularly the open-ended responses, provided enough detail about these universities' reactions to the NRC study. In general, reactions fall into the categories discussed above, primarily including a hesitancy to use the study results due to the age of the data and questions of accuracy.

Final Case Selection

Prior to the final selection of cases, a second analysis of the prospective pool of case study universities was conducted based on a review of wholistic institutional responses. Differences between the two levels of responses – chief academic officers and graduate programs – were reviewed and highlighted. Specifically, it was determined that the selection of cases should focus on the similarities and

differences between the central administration level responses and the program level responses regarding reported use of the NRC study results. Only the 22 institutions where the central administration respondent agreed to the case study were included in this final analysis. Because of the level of heterogeneity in program-level responses, some universities could legitimately be placed in more than one quadrant.

The matrix in Table 5.3 shows the responses on use of the NRC study by level.

Table 5.3 – Matrix of Case Study Possibilities Based on Use and Level

	Use of NRC Study (Central Admin)	**No Use of NRC Study (Central Admin)**
Use of NRC Study (Programs)	Case Study #1 (7)	Case Study #2 (4)
No Use of NRC Study (Programs)	Case Study #3 (6)	No Case Study (6)
Total	*12*	*10*

Note: totals do not sum equally due to response rates on individual use questions and because some universities may fall into more than one quadrant based on heterogeneous program-level responses.

Case study #1 would be a university with alignment between the central administration and programs reporting use of the NRC study outcomes. Case study #2 would be a university with different responses between the two survey respondent levels with the central administration reporting no use or planned use of the NRC study results yet the program-level responses reporting use of the study results. This outcome could lead observers to wonder why programs might use the results if it not valued on their campuses or among their leadership. Some

perceived value at the program level must have been present. Case study #3 would be a university with the opposite dichotomy in that the central administrator respondent reported use of the NRC study results, yet the majority of the program-level responses did not. Similarly, this discrepancy was reviewed further in the case study interviews. As in the first analysis, the fourth quadrant – where both the central administration and the programs reported not using the NRC study results – was deemed unnecessary for further in-depth study as part of the dissertation.

Based on these analyses, the 22 institutions where the chief academic officer had agreed to cases were reviewed across the four quadrants from both analyses. There were seven institutions with no program-level survey respondents, which eliminated them from consideration as case study sites. Many of the institutions fell into the same quadrants on both analyses described above, ensuring the perceived goal achievement of the NRC study and reported usage based on university or program level were both taken into consideration at the selected sites.

Seven institutions rose to the surface as ideal candidates for case studies. They were ranked based on the number of programs present at each campus and the level of participation in the program surveys, as well as reviewed on other characteristics such as public or private status and research classification. Quadrants 1 and 3 had a primary university candidate as a case study site with one backup institution, and quadrant 2 had a primary university candidate with two backup institutions.

Three site visit case study universities were initially selected for further study and site visits. Invitations were first sent to the three chief academic officer

respondents during the early fall 2012 validating their agreement to participate as confidential case studies and ensuring their level of comfort with their campus' participation. Each of the three agreed to participate, so communications with the backup institutions were not required. Once agreement to have each campus participate was secured, invitations were sent to pertinent administrators and graduate program directors, almost all of whom were also survey respondents. The site visits all occurred during the fall and winter of 2012.

The identity of the three case study sites will be kept confidential. As part of the IRB consent form process, the interviewees were made aware of what information was to be included in the final dissertation (e.g., public or private status, geographic location) and what information was to remain masked (e.g., university and interviewee names, exact degree program names, and any identifying campus descriptors such as titles or programs).

Case Study Themes

As part of the case study process, possible areas of inquiry for each participant in their interviews include the themes in Table 5.4 below. Robert Weiss' book, *Learning From Strangers: The Art and Method of Qualitative Interview Studies*, served as the basis for planning the process for identifying respondents, selecting the interview topics, conducting the interviews, and analyzing the results. The book details the rationale used for less structured interviews in order to obtain a thorough understanding of a series of events, specifically to compile and use for case studies.

While a script of questions and topics was developed and distributed to interviewees prior to each visit, time was allowed for the discussion to diverge off-script to enable capturing more detailed information from the interviewee. These themes, together with all appropriate consent questions and information, were included in the IRB submission for review and approval.

Table 5.4 – Case Study Interview Themes

Theme	Description and Questions
1	Ask respondents to provide background on the process utilized for collecting and validating the data submitted for the NCR study in 2005-06. Most likely this background will show how robust centrally-administered data resources are on a campus. Institutional research offices will likely be the most valuable resource for an overview and assessment of this preliminary time period. What policies and best practices were used?
2	Ask questions that delve into the preparations the university undertook prior to the NRC study release in 2010, such as conferences with campus faculty, preparatory workshops, and media briefings or public relations tool creation (e.g., websites, press releases). What policies were impacted during this time period?
3	Ask questions that allow for description and assessment on how the university and programs reacted to the NRC study release, such as press releases, executive summaries, briefings to senior administrators, faculty workshops or presentations, email notifications, or news stories on campus and locally.
4	Ask questions that allow respondents to describe and assess university processes used after the NRC study release to explain the results to the campus beyond the immediate reactions to the rankings. For example, were there meetings with participating programs to discuss their results, how was the data collection of errors handled, and what types of central or individual program reports were created using the NRC study data or rankings.
5	Allow for further discussion on how the university and programs used or plan to use the NRC data, including the rankings and the individual data variables. What types of activities have they done or policies have been changed with the results to improve the quality of programs on campus? Are the results being used for more evaluative purposes or more action-oriented purposes? What is impeding the ability to use the database and results even further on campus?

6	Provide respondents with information about their campus responses at both the central administration level and the program level as well as comparisons to all program responses combined. Ask respondents about any similarities or differences in their reactions both on campus and nationally in their field. For example, if the chief academic officer reported higher rates of use on campus or expressed more agreement about the influence of the NRC study on the quality of graduate education on campus than did the program chairs, why might that be?
7	Ask the respondents to assess the value of the NRC study for their campuses, including cost/benefit questions, resources expended, validation of the data, future oriented plans for resources and data collection, and how their campus assessment and quality processes may have changed as a result of the NRC study experience. Seek out any public policy changes or recommendations that might be nationally applicable.
8	Provide an assessment of how the NRC study has elevated the discussion about graduate education and program quality on their campuses. Most campuses will relate the NRC study through program review activities, so allow time to discuss their review process and how it might have been improved due to the NRC data and process generally. Seek out any public policy changes or recommendations that might be nationally applicable.
9	Provide information on the dichotomy presented in the survey results concerning use of the NRC to persuade decisions/change versus justify decisions/change. Ask respondents to assess this finding within the context of their own campus activities.
10	Generate discussion about the respondents' perceptions of institutional factors that determined how responsive they were to the NRC study. For example, how does leadership's role in change and the perception of the study impact use? Determine across cases if there are regularities across campuses that led to enacting change and use of the study results.
11	Ask questions that allow respondents to transcend their campus uses and responses to national efforts. Do rankings force conformity on programs such that they drive out experimentation that can in turn improve quality?
12	Ask respondents for their prediction if the NRC study will occur again, and what steps, if any, their campus is taking to prepare for this next phase. Other assessment activities occurring on the campus due to accreditation, state or regional studies, and quality improvement efforts will be discussed. Seek out any public policy changes or recommendations that might be nationally applicable.

Case Study Findings

All case study site visits were completed by mid-December 2012. The information gathered was assessed and reviewed for shared experiences and best practices for data use and assessment activities. Of particular emphasis were any institutional factors that overlap and might have contributed to use of the NRC study and its results.

The three cases have been compared and contrasted where appropriate to identify findings that indicate trends or potential best practices that lead to institutional and program level changes. The dissertation results focus on findings specific to program and quality improvement that have public policy implications and could be relevant nationally. Specifically, each campus' use of the NRC study as an information instrument for enacting change and continuous improvement was analyzed. While one university's change processes cannot be generalized, the comparisons provide insights into the policy implications for institutional change as a result of similar quality or rankings studies.

Overview

The three case study universities will be kept confidential in the dissertation, and no identifying information will be reported. Two of the universities were public institutions and one was a private institution. Geographically, one institution each was located in the Northeast, South Atlantic, and South Central regions of the country. One institution was a member of the AAU, two were members of the APLU, and all three were very high research activity campuses in the Carnegie classification system.

Overall student enrollments were greater than 12,000 students at all three campuses, making their profile large four-year campuses. Each case study site had at least three of the six graduate programs represented in the dissertation's program-level surveys, showing a breadth of graduate education opportunities and research foci at each campus.

Interview requests to the groups described above were met with a willingness to share time and information in virtually all cases. Across all three campuses, only one survey respondent program and one program director who did not respond to the survey were unwilling to meet during the on-site visit. The Provost and associated central administrators were willing to meet at each campus, including graduate deans, institutional research directors, and other academic affairs representatives in Provost's offices such as senior vice/associate provosts responsible for relevant areas like research and faculty support and development. All case study interviews occurred in person except for one program director interview that occurred via telephone conference due to travel schedule overlaps.

The overall sentiments from the case study interviews mirrored the survey findings. The NRC study was generally perceived as a flawed study, in large part because of the data issues and the long length of time between data collection and publication of the results. With a few exceptions, interviewees expressed almost a sense of remorse for feeling such sentiment. Hopes were high for the NRC study. Faculty and administrators wanted to use and rely on it for substantive discussion and improvements on their campuses related to graduate education. Interviewees

found it unfortunate and extremely disappointing that the study could not meet those expectations and that so much time and resources were wasted.

At the same time, there was some perceived value achieved by the NRC study. Without exception, each interviewee relayed that the NRC study allowed a spotlight to shine on their graduate programs and the overall research endeavor at their campuses. Some found it more beneficial than others, and some maintained that such a focus was already present at their campus even without the NRC study. Though similar, the reactions on each campus slightly differed to this line of questioning. Some felt their campus already highly valued graduate education, and the NRC study only served as another data point in the long-standing discussion about research and graduate student training. Others felt the NRC study allowed a renewed emphasis on graduate education, often dwarfed in campus debate by more pressing, and resource-intensive, undergraduate instruction and enrollment growth needs. But all agreed that the NRC enabled their campus to pay attention to their graduate programs and engage in further discussion about quality, program improvement, and future needs.

The next sections provide a summary of each of the three case study institutions and their reactions to the NRC study on their campuses. Following those summaries is a compilation of overlapping best practices and national recommendations related to data use and assessment practices, including their implications on public policy.

Case Study: Quadrant 1

Case Study 1 is a public institution located in the South Atlantic region. Both the central administration survey respondent and the graduate program respondents generally agreed that the NRC study and its results were in use on their campus, both centrally and at program levels. The central administration survey respondent believed the NRC met its goals and reported the campus was using the NRC study results on their campus, although the extent to which they are truly able to use results is somewhat variable across programs. In interviews, central administrators articulated the value of rankings data and studies such as the NRC and appreciated the need to use data intelligently for their assessment, accreditation, and accountability activities on campus. They could be characterized as troubled by the fact that the NRC study did not facilitate further use and widespread awareness, primarily because of its complexity. Due to a greater emphasis on accountability in higher education at all levels, this campus understands the need to use statistics and metrics more heavily. At the same time, they also know there must be tangible and measurable outcomes in use lest they risk alienating the campus and failing to meet those same accountability standards.

The campus administration routinely uses scorecards, dashboards, and metrics for goal-setting on campus and with external constituent groups, including the use of data on peer campuses and programs and aspirational peers. Graduate education is one core area with some key metrics, though interviewees acknowledged it enjoys less of a public focus than undergraduate matters. Time and again, interviews uncovered a desire for simple and understandable measures

to track successes and improve themselves in core areas related to graduate education. Operational definitions in graduate education are quite varied across disciplines and campuses, a fact that hampered further use on this campus. Without common graduate-level data on which to build and compare themselves, this campus tends to reflect more internally and focus more heavily on undergraduate-centric metrics, which do have simpler and accepted common definitions. Accrediting bodies have taken some steps toward identifying a common threshold and highlighting requirements for assessment and student learning outcomes, but most interviewees suggested that was not sufficient. Support existed for national efforts to help define core graduate education metrics with common definitions upon which all programs can be measured.

In general the programs at this institution performed well in the NRC study. The program directors interviewed believed the study showed successes and growth over time, generally even surpassing their reputations in their respective fields. This sentiment was echoed by the central administrators who were interviewed as well. The graduate program faculty are seeking ways to spotlight their strengths, both to improve themselves internally but also to prove their quality to others – prospective graduate students, faculty hires, and their deans' offices. Comparative peer data would be useful to programs under both scenarios.

Programs want accurate data on which to base decisions and compare themselves to peers, especially those perceived as higher quality or aspirational peers. At the program level faculty understood their Provost's push toward more data-oriented decision-making on campus. Some programs are using data more

extensively than others, but the emphasis on doctoral program assessment was obvious in every interview. One program interviewed, in particular, takes data very seriously while looking for improvement opportunities within their department. They highlight the NRC study results in their web materials and encourage continuous faculty and graduate student discussion about the measures. Their recent program review and self-study experience placed the NRC study findings and metrics at the forefront. They reported that the study helped them develop appropriate goals and reconsider who their true aspirant peers might be given their quality rankings on several measures had grown increasingly stronger since the last NRC study. In a sense, they were stronger than even they realized and grasped the opportunity to enact changes to continue to improve themselves. They firmly believe the NRC will contribute to quality enhancements in their graduate program and by extension on the campus.

The central administrators interviewed likewise expressed a willingness to support quality programs and improvement efforts with some resources, even in limited budgetary times. These types of financial incentives were viewed as critical for putting forth time and attention on data collection and analysis projects, especially in programs where such efforts were not part of existing culture. Doing well in studies such as the NRC and seeing tangible benefits as a result, many said, makes everyone want to improve and do even better.

Interviewed program chairs and central administrators were clear that they are looking for ways to publicize their successes as a campus. They believe the quality of their educational and research offerings has improved in recent years,

even the past several decades, but because reputations lag in their disciplinary communities and across the country, their programs may be held back unnecessarily from garnering top students, faculty, and competitive grants and awards. A sentiment expressed by many on campus was that the true value of studies such as the NRC can be muted by the lack of visibility and limited use opportunities. If some of this campus' recommended changes to study methodologies and usability factors could be implemented, they believe the rankings data and study results would evaluate them more accurately and they would thus enjoy tangible benefits on campus and nationally from their efforts to improve themselves.

Case Study: Quadrant 2

Case Study 2 is a private institution located in the South Central region. The central administration survey respondent reported the NRC study and its results were not in use on the campus, yet the graduate program respondents generally agreed that the NRC results were in use in their programs. In the interview, the Provost expressed a common opinion among peers that the study is not readily usable nor will it be repeated with the current format. The sentiment was that the higher education community lost faith in the NRC study due to the delay and the complications of its release and methodologies. This campus is already very data-driven with key central resources dedicated to ongoing accreditation and assessment activities. They supported the campus response to the NRC centrally, generally making it easier on individual units to participate. Although interviewees agreed with the line of questioning about the NRC highlighting graduate education

on campuses and nationally, it was apparent that each interviewee already believed graduate education was a focal point on this campus.

Rather than attempting to use the NRC study data centrally, this campus has shifted toward third party sources for data about their faculty and students. It also enjoys a very robust institutional research office to support routine data reporting efforts and special projects, including at least two recent task forces to review graduate education on campus. The campus administration wants to be able to manipulate data for their own purposes and a variety of projects, which was not entirely feasible with the NRC study data. One recent example of changes using their own data efforts included planned reviews of emerging research areas, which may obtain central resource investments to ensure targeted and sustained growth. Also, their graduate college was restructured based on data from campus feedback and an analysis of needs.

Specifically this campus wants to review comparable data across campuses, including peers and aspirants, incorporating baseline data and an ability to track trends over time as they make improvements and changes to policies on campus. Their program review process is non-standard, but discussions were ongoing about how it may be revised now that additional sources of data are available to the campus. Many of these efforts are possible because of their access to external data sources and internal infrastructure for reporting and data analysis. Several interviewees offered suggestions for the types of comparative data that would be valuable to have nationally. For such efforts to be successful, they also assumed

common definitions would be developed and a coordinated effort to include all types of graduate programs and institutions would be in place.

The programs at this campus did quite well in the various rankings methods in the NRC study. Several central administration interviewees indicated the study simply reaffirmed that they were doing well. That reaction, together with the pushback nationally to the study's data issues and delays, enabled them to have a minimal reaction to the study's results at the campus level without any desire for centralized coordination of its use. However, several of the programs interviewed did indicate that the results had been incorporated into their program's activities. A few mention the results in their web materials. The complexities of the ranges of rankings made them difficult to use; however, the individual data elements in the NRC study allowed for cross-campus comparisons. Several programs reported breaking down the study to review those characteristics that were important to them. They used them to highlight their areas of strength and seek opportunities for improvement, though they also reported in interviews that the results did not necessarily persuade them to enact substantive changes.

For example, one program was in the midst of a restructuring effort when the results were released. The NRC study results were used as a tool to motivate faculty to engage in the effort and really understand their overall strengths. They were surprised, somewhat disappointed even, that their administration did not appear to value the results, as they found important information buried within the study. As part of the restructuring, this particular department made their weaker areas a point of emphasis. How could they better use their resources, new faculty

hires, and time to improve themselves in these areas? What opportunities might they have to strengthen key areas within their program to attract even better students and faculty? Although they felt they were a strong and adaptive unit already, the NRC study served as a tool for program debates on areas of scholarship and internal departmental policies. The program felt the NRC study persuaded them to implement some changes, though the program did not report they experienced substantial change or an increase in the quality of their program solely due to the NRC study. For them, graduate education and research were already points of emphasis, quite strong in nature, and highly valued on their campus. Even if it was serendipitous in nature, the NRC study did serve as a valuable set of information about their standing to validate their strengths and provide them a chance to look for development opportunities.

Several interviewees brought up the fact that the campus and programs have begun to revise their focus toward more types of qualitative and anecdotal feedback, especially from graduate students. They are also beginning to prioritize addressing some new trends in graduate education, including faculty mentoring efforts, career alternatives, professional development initiatives, and alumni tracking. This campus has established the infrastructure necessary to be a data-oriented, decision-making campus and is gradually moving beyond the focus on data, in large part because it appears to be part of their accepted culture. This characterization and shift in direction does not mean this campus is going to cease using data; far from it. Rather, they are comfortable with their tool set and their overall university standing, or at least their awareness of their standing, and are ready to use those resources

and knowledge to embrace some of the larger issues facing graduate education today.

Case Study: Quadrant 3

Case Study 3 is a public institution located in the Northeast region. The central administration survey respondent agreed that the NRC study and its results were in use on the campus, yet the graduate program respondents generally reported that the NRC results were not being used in their programs. The Provost and other central administration interviewees focused on the value the NRC study and results had in their campus doctoral program review, a special study implemented around the time the NRC results were released intending to assess thoroughly each doctoral program on campus. They expected the NRC study to provide substantial data for this effort, and the underlying metrics did prove useful for this project by offering comparative data and national norms on key variables. The NRC ranges of rankings, however, were not deemed valuable even though many of the programs on this campus performed well in the various rankings methodologies.

Similar to the Case Study 2 institution, this campus already also relies on third party providers for data about their research endeavor and faculty productivity. These data, together with the campuswide doctoral program study, were perceived among the central administrators as some of the best initiatives ever undertaken on the campus. Programs were asked to review their own faculty, student, and research data and compare themselves to national norms. Peer programs were reviewed and aspirants were identified as comparative programs against which they could learn. The administrators interviewed recognized that all databases and

metrics can have limitations and inaccuracies, but by placing much of the debate and discussion in the hands of the faculty, they felt the data would ultimately help guide quality improvements. Programs were asked to face their weaker areas and identify ways to address them. Ultimately all campus doctoral programs are expected to be rated in a classification system with identification of emerging areas for investment, and subsequently some areas for disinvestment.

Although the doctoral program review process was led by a highly-respected faculty member and stressed an open and transparent process, some programs expressed a bit of hesitation about the usefulness of the massive exercise. They did not necessarily realize that some of the data they were reviewing came from the NRC study, which can partially account for the survey responses that the NRC study and results were not in use in their graduate programs. Some reluctance may always be present in such reviews because of the uncertainties in how a program will fare and the associated consequences. It should be acknowledged that the specific programs interviewed as part of this study and the timing of the site visit for this dissertation – after the initial campus reports were written and evaluated but before any tangible actions had been announced – could have contributed to these perceptions as well.

That said, with one steadfast exception, most programs did see some value in the self study and data comparison exercises. They noted that the faculty always want to improve their programs to ensure competitiveness for graduate students, research grants, and faculty hires. If nothing else, because the administration clearly valued data-oriented information, programs felt they could use such studies

and results as part of their lobbying efforts for additional resources or policy changes affecting their programs. Every program expressed a desire to have a good showing and valued the transparency of the doctoral program review process. One program even indicated that they wished they had more time to spend on identifying ways to use data to improve their program rather than spending time reviewing, correcting, and explaining the data. In contrast to the institution in Case Study 2, the infrastructure for data reporting and analysis is not fully in place or as robust on this campus, although they are clearly moving in this direction. Thus the programs have not yet achieved a steady state of use for the intended purpose of quality improvement.

The exception voice from the interviews merits mention. This program director expressed concerns for abuse of data – not intentionally but because of the complexities of the graduate research and training enterprise. If the data are not accurate, then such large studies are not worth the time spent on them. Additionally, this interviewee pointed out that a key flaw in such studies and data sets is their treatment of interdisciplinary programs, a point raised in some survey responses and at the NRC Convocation. Because faculty lines, students, and resources are structured differently in certain disciplines, identifying how to count and rate faculty productivity, research, teaching, and graduate student mentoring can be virtually impossible. It may be too difficult to attempt such studies nationally because of these inherent issues, at which point, this interviewee would argue, the study is no longer worth the time and energy devoted to it. This viewpoint is not uncommon

among faculty and would need to be acknowledged and addressed if any national efforts such as the NRC study were to continue.

Similar to the interviews at other case study campuses, this institution benefitted from the campuswide discussions about graduate education. There were retreats and open meetings to discuss the doctoral program reviews and determine goals for the campus. The attention and time devoted to graduate education was welcomed, especially in tight budget times. While still on the cusp of seeing the results, many appreciated the opportunity to have a say in the future planning and direction of the campus. It was critical, many interviewees indicated, that faculty support these types of processes. They need both to understand the data and also contextualize it by shaping their own program's story. The data alone are not sufficient for appropriate decision-making.

Again echoing the feedback heard from the other two case study sites, national norms and common data definitions would prove enormously beneficial for studies such as the one this campus is undertaking. Several national entities were recommended as possible leaders and coordinators for this effort, including CGS, the National Opinion Research Center at the University of Chicago (NORC), or the Association for Institutional Research (AIR). Many acknowledged the difficulties in this type of undertaking, if for no other reason than disciplinary differences would be difficult to overcome on many metrics regarding student quality, resources, and funding sources. Yet, this campus also shows it is possible to learn and improve itself through review processes such as the one they undertook. Thus it appears

more can be done to encourage such types of growth and quality improvements at a national level.

Limitations

Several limitations exist in the design as laid out in this chapter. As mentioned above, the case studies and findings, while instructive, are not generalizable to other universities or other rankings or quality studies. One experience at a university is particular to its own individuals, resources, and circumstances. At the same time, one goal of a case study is to provide a detailed look at a particular experience to enable some lessons and sharing to grow out of it. The dissertation's goal was to identify overlapping experiences at the universities studied and provide other campuses a roadmap to implement sustainable change and policies that influence quality improvement in their graduate programs.

Another limitation involves the selection of the case studies. While care was taken to select appropriate universities to study further, no single university replied to every survey and agreed to serve as a case study in all responses. Thus, negotiation was required to secure participation at each case study site, which did not always result in interviews with every desired program director or administrator. The number of non-participating program was fortunately limited to two across all three case study institutions. Additional cautions in the findings involve acknowledging that programs may not have been fully comfortable discussing issues of quality when they may diverge from their campus leadership's perceptions and vice versa.

Finally, as described in the survey response chapter, there is skepticism and cynicism surrounding the NRC study. Obtaining agreement to be a case study from all participants at a campus was difficult when some felt the NRC study was inherently flawed and unworthy of use or further study. Many of the interviews diverged into data use and assessment topics beyond the NRC study alone, which while still meaningful, was not the sole focus of the interview. These broader opinions, however, did lead to many of the best practices and policy recommendations described in the final chapter.

CHAPTER 6 – CONCLUSIONS

This dissertation has described impacts on assessment activities and institutional behavior, and informed policy impacts, of national rankings and assessment studies such as the NRC, specifically on the overall quality landscape of higher education at the graduate education level. While it is difficult to attribute change to one study or set of data, one can draw some conclusions and make policy recommendations for incorporating faculty, student, and program level quality data into an overarching assessment and planning process on campuses.

The results and analyses of both the surveys and case studies inform this dissertation's findings and policy recommendations. Of particular note is the impact of the NRC study on higher education institutions, both central to the university and at the graduate program level. The attention and focus on graduate education has been beneficial to the community, engaged faculty, and has led to broader discussions about the needs of graduate training to ensure the United States retains its status as the preeminent provider of doctoral education. Both private and public universities can benefit from focused study on quality and program improvement. Public universities will generally experience increased pushes for change in part due to the external pressures on affordability and governmental accountability pressures.

Returning to Gormley and Weimer's work, the findings presented in earlier chapters do show that data from quality studies can be used to make policy

decisions and enact change. This dissertation's research confirms many of the theoretical concepts described in Chapter 2. Quality and performance indicators can be used as policy tools in decision-making processes at universities. A variety of responses were reviewed via the survey results and at the case study universities, specifically among the managers and overseers on campuses who have discretionary oversight for policy priorities and implementation both at the campus and program level. Gormley and Weimer argue for the use of report cards and rankings data to frame policy issues and facilitate end users' ability to evaluate the results and determine their own avenues for collective action. Such outcomes are difficult, yet the NRC study did lend itself to several like strands of use, such as the all-encompassing doctoral program review in Case Study 3 or the detailed program evaluation experiences in some of the departmental and graduate program efforts.

The culture within an organization is critical to the level of engagement in change processes, a theme that will be detailed further below in the specific best practices. Gormley and Weimer's model of organizational response, particularly the desire to gain functional responses to report cards and data, stresses the importance of focused attention and mission-driven activities on process improvements. Cultural responses can generate feedback loops related to resource use, discretionary change, and flexible actions. If the environment is supportive of quality outcomes and improvement efforts – as was seen in the case studies highlighted in this study and many of the survey responses implying functional responses – then the overall mission and organizational culture can be tailored toward sustainable change.

For transformational and sustainable change to occur, key factors will be in place – themes addressed in each of the core theoretical models that drove this dissertation's research. As Feldman, et al. and Aldrich and Ruef argue, the environment must be open to change, appropriate leadership and focus will be present, incentives and understanding for change processes will be apparent to campus constituencies, comparative data should be available and in use, and new routines need to be established that evolve organizational culture – all for organizations to be effective "academic learning organizations" (Dill, 1999, page 128). These themes were identified in the survey results and case studies, especially when considering the differences between the central administration results and the program-level results. While each theoretical element alone was not sufficient for impacting change processes, in particular at the case study universities, they were necessary for such activities to occur and have broad campus support.

Even with external pressures present – such as rankings studies, public and governmental accountability standards, and performance expectations – substantive change is difficult at both public and private universities. It is, however, possible with key factors and best practices in place. Data collection and analysis processes and heightened focus and attention about graduate education can be positive outcomes from participation in quality rankings studies such as the NRC.

This chapter proceeds with an analysis of these best practices and recommendations compiled from the research. The chapter also considers the public policy implications for this research and concludes with some summarized thoughts about the findings and suggested areas for further research.

Best Practices

This dissertation intends to identify findings specific to program and quality improvement that have public policy implications and could be relevant nationally. Both the survey results and the situations at each of the three case study institutions are unique to the campus' own circumstances, leadership priorities, and resources. However, this research did uncover multiple areas of overlap in priorities and policies related to data use and assessment, which can be identified as best practices that lead to institutional or program level changes. These findings are consistent with Gormley and Weimer's theories regarding the influence assessment data can have in creating quantitative analyses and framing issues to provide detailed information for policymakers.

Best practices, identified by survey and case study results and informed by theory, are broken into three categories: institutional structure and culture, data recommendations, and reframing quality studies.

Institutional Structure and Culture

At all three campus case studies, the central administration's support for using data as part of the decision-making process was crucial. In today's age of higher education accountability, most institutions do not view this as a choice; however, the extent that quantitative metrics and peer comparisons are in use on a campus varies widely, especially at the graduate education level. Having dedicated resources to work on these activities – whether in the Provost's office or in central functional offices like a graduate college or an institutional research office – benefits the campus as well. Then faculty and staff with other priorities can be supporters for

the data efforts instead of finding themselves as the far-too-strapped leaders for such efforts. Reports are increasing of universities hiring staff dedicated to collecting and analyzing data, which can be positive developments indicating campuses value strategic growth and improvements – assuming the positions are not intended to look for opportunities to game or play to the rankings (Trounson, March 2013).

At the same time, faculty will necessarily play a key role. They have a vested interest in the quality and state of their departments and graduate programs. The campus culture will enable an emphasis on embracing change at all levels and not have strategic change efforts perceived solely as administrative bureaucracy. If the campus leadership places high value on such activities – to the point that larger studies are occurring or in use on a campus, discussions occur surrounding future quality enhancements, and campus culture begins to embrace the outcomes of such efforts -- the campus will engage in these efforts.

As such, the various constituencies on a campus or in a graduate program will appreciate and take interest in these quality review studies, including faculty, administrators, and students. This does not necessarily mean that everyone on campus must agree. As seen in Case Study 3, there can be healthy, dissenting opinions that ultimately could aid in creating more inclusive and accurate data use policies. But campus groups and constituencies cannot fear the actions to be taken as a result of studying and applying data to the decision-making process. Openness, transparency, and broad ground-level input were critical elements in any campus or program description of using the NRC study. To engage with large

amounts of data and use it to assess possible changes and future directions requires substantial amounts of time and energy. These expectations should be apparent. To make it worthwhile and ultimately successful, all groups should ideally be engaged and supportive of the process.

One way to foster that engagement is if participants see tangible links between the study processes, use of data, and any potential outcomes. Budget changes or resource allocations are the most common outcomes that faculty and students may expect to receive toward implementing quality improvements. If there are tangible outcomes on campus that make change worthwhile, groups will engage. In his work reviewing possible cost containment strategies for higher education, Massy (2013) shows how process audits can aid universities in understanding how the use of resources can be linked to outcomes and the adoption of best practices on campuses. With better data and information about quality, process and resource improvements on campus can and would occur.

Financial incentives need not be the only outcome. Rather, multiple areas of use for the NRC study were identified in the survey results and further discussed in the case studies. Program reviews, policy revisions, student and faculty recruitment, and public relations activities were all addressed as effective uses for quality and rankings data, many of which would have tangible outcomes on campus and program-level policies, recruitment efforts for faculty and staff, and resource allocations. Once focus and attention on these activities can become institutionalized, and data and quality studies accepted as information instruments in a campus' culture, the outcomes of these activities – higher quality hires and

students, improving disciplinary reputations, or greater awareness of campus strengths with external boards and governmental groups – could prove to be the ultimate benefit.

Data Recommendations

For these efforts to become reality at the campus level, there is a need for a national database or common definition data set for graduate education. Attempts are underway to improve the types of outcomes data available to measure higher education performance, though they focus more heavily on undergraduate and community college indicators (Fain, 2012). There are also international models to follow, such as the Australian and European Union tools used for monitoring universities and comparing research productivity and student learning outcomes (Trounson, February 2013). A core taxonomy of disciplines, including the professionally-oriented doctoral programs that were excluded from the most recent NRC study, will need to be established for true comparisons, perhaps taking more consideration of the Biglan Model dimensions into account (Biglan, 1973) as described in Chapter 1.

At the graduate level, core factors such as admission statistics and yield rates, time to degree, student funding and stipend levels, teaching and research skill evaluations, research and publication activity, and other student support services do not currently have common definitions across disciplines or campuses. There are efforts among the AAU campuses to identify common data elements and exchange data among themselves for comparative purposes (AAU Institutional Data Committee Memo, 2012 also reiterated in spring 2013), but several case study

interviews – even those at the one case study site which is a member of the AAU – indicated such efforts are not yet broad enough to encompass all types of graduate training and university missions.

There are also known cautions within the graduate community about some of the shared datasets already in place (Gater, 2003). A data set at the graduate level needs to be common across all graduate-level institutions to be most effective and representative, perhaps beginning at the doctoral level but eventually moving to encompass master's level programs and students as well. Calls have begun for such work to occur (Olds and Robertson, 2012), though admittedly it will not occur quickly or easily. Gormley and Weimer's work argues that data can play a role in policy framing leading ultimately to better decision-making, a core theme of public policy and one seen throughout the case study and survey response experiences detailing functional responses to the NRC study results.

Stakeholders must be engaged continuously to enable user-driven rankings and usable data (van Vught and Ziegele, 2012). Accommodations for any shortfalls related to institution-reported data and the unique interpretation of key metrics will need to be taken into consideration. Key members of the graduate education community and national leadership will be called upon to assist with framing the issues and forming policy recommendations in this area, which will require high levels of engagement and time. The successful process will be iterative and inclusive, enhancing buy-in but also increasing the efforts' time table and complexity. The lessons learned from problems identified during the most recent NRC study can remain in the forefront to ensure they are not repeated.

Optimistically, once the variables and metrics are settled on and defined, the data will need to be updated on a routine basis through mechanisms that facilitate university responses. The community cannot wait 10 or more years between releases, as was the case with the most recent NRC study, because of lost institutional memory, the time and effort spent to ramp up for such broad studies, and the overwhelming nature of this type of study. If key metrics are identified, campuses can tailor their data infrastructure, including key staff and resources, to report their results. As reporting becomes more routine, trend data will naturally result and prove valuable to the campuses and for national research on graduate education.

Comparative data using peer campuses and programs, as well as peer aspirants, is critical for quality improvement efforts. Programs need to be able to hone in on key metrics that matter to them in their disciplines, even if explicit rankings are not the end result. Some programs may also value certain characteristics over others, and a robust data set inclusive of all key variables important to graduate education assessment should be minable depending on need and preferences. The variables mentioned above were discussed in several interviews, and it appears every discipline could likely find value in at least some of the measures. If engaged stakeholders perceive value from the metrics being requested, the data set has much more likelihood for usefulness and longevity.

As previously described, a national organization would ideally take the lead role in coordinating this effort. Academic studies have also suggested a mixed role of governmental support with an emphasis toward impartial, non-media outlets

(Hazelkorn, 2007 and 2008). Possibilities include CGS, NORC, or AIR with heavy participation from individual campuses. Theory, specifically Gormley and Weimer's and Aldrich and Ruef's work, shows that engagement among multiple levels should be present to ensure buy-in and strong trajectories toward program improvement and change. Universities themselves can be actively involved to assure their needs and outcomes will be met.

No ill will toward the National Research Council was intended when interviewees made suggestions for other coordinators of these data efforts at a national level. The National Academies and the National Research Council still have a role to play by remaining involved in centralized data collection efforts, especially because the furthering of graduate education is critical to their mission. They remain engaged, and should remain so, in efforts to strengthen the country's research enterprise and lobby for funding for research institutions (National Academies and NRC, 2012). The AAU continues to stress the importance of graduate education and the public research enterprise (Rawlings, 2012). But interviewees believed it was time for another organization to take the lead coordinator role for a common data set at the graduate education level and to enable the NRC to focus on the broader goals of maintaining support and quality for US institutions of higher education.

Similarly, a shift away from the massive decennial study toward one of routine metric reporting will be useful on many levels and may be better-suited to one of these other organizations who have such routine data collection and analysis efforts already part of their mission. In her comparison of a series of rankings data efforts,

including the NRC study, Hicks (2008) suggested that continuous reviews of the methodologies have the potential to lead to better improvements on campuses and to avoid attempts to tweak or game the ratings systems. As a collective effort, drawing upon the expertise and staff resources at these various organizations and relying on input from the faculty and administrators engaged in graduate education, movement toward a centralized, comparative, common data set appears more likely to succeed and have the support of the community.

Reframing Quality Studies

The purpose of rankings and quality studies such as the NRC needs to be reframed. In many of the case study interviews and in some survey comments, there was a desire expressed that the higher education community should move beyond rankings and publicity on who might be number one toward enabling program improvements on key characteristics. The shift in purpose embraces the use of data as true information instruments leading to change, which was one of the stated hopes of the most recent iteration of the NRC study. However, the faults of this study, especially the time lag and the complexity of the study methodology, hindered its acceptance broadly among the graduate education community. The time and focused attention needed to truly engage in such efforts also contribute to the complexities in moving forward on this idea, both for the NRC results and more generally.

Some inroads and best practices have followed the NRC study release but more can be done, especially at the graduate education level (Dill and Beerkens, 2010; van Vught and Ziegele, 2012). For example, in September 2010, the graduate

education-focused Strategic Leaders Global Summit – sponsored by the US-based CGS and the Australian Group of Eight – developed 10 principles for measuring quality in graduate education, including a shift in focus toward one of improving the quality of training and student learning, renewed emphasis on internal and external review processes, and refining quantitative and qualitative tools and methodologies for measuring quality (Council of Graduate Schools, September 15, 2010).

As another example, the Australian efforts discussed in the previous section aimed to achieve such goals, including tools using publicly available data to chart university activity on broader measures – such as teaching, student data, and knowledge exchange data – than has been possible previously (Trounson, February 2013). They endeavor to have each institution embrace what they value and perform well in, acknowledging that not every characteristic can be top-rated at each university. Institutions should be "free to build on their strengths" instead of chasing unrealistic end goals of prestige (page 2).

Similarly, recent debates have occurred within the graduate education community about the value of low-ranked programs. Cassuto (2013) has engaged in a series of articles calling for right-sizing graduate admissions, recognizing the demand for faculty nationally and ensuring that the scope of graduate programs only reaches to fill those gaps. It can be argued that good work among scholars can occur at any level and that the emphasis on rankings is simply to compare what one is already looking to compare instead of enabling graduate programs and students to set their own types of training agendas and professional skills. So-called lower-

ranked departments should instead focus on what they do well instead of being as concerned with what goes on above them.

Selingo (2013) echoes these arguments by harkening back to the prestige-seeking discussions earlier in this dissertation. Institutions are in a losing battle if they focus only on getting ahead in the rankings, he claims, instead of experimenting with innovative educational opportunities, which will likely never be reflected in rankings methodologies. By focusing on the good work programs are doing, not to mention the students they are preparing to do well in the types of jobs in which their program can play a large role, universities are instead too often focused on prestige-seeking activities in a zero-sum game. This debate centers on the fact that there is not time for such activities, especially in difficult budget climates, and each program should instead embrace what they do well to remain relevant.

These themes are also consistent with Dill's remarks (2011) on university reforms and change potential. Governmental influences and reforms are not necessarily leading to institutional diversity, rather, toward institutional imitation. At the graduate education level, the drive for more resources, research, and faculty productivity gains – driven in large part by reputation-seeking faculty and university administrators – contributes to homogeneity among campuses. Rather, with an emphasis on program improvement and student learning outcomes measurement and growth, campuses and faculty could instead focus on diverse missions and working to fulfill their own particular niche in student training. Dill highlights the NRC study as one with the potential to influence this type of change (page 9).

The task for graduate education in the United States is complicated with many steps remaining. But the end result will be one that distances itself from simple overall rankings, which no one believes are fully accurate or useful, and moves toward a common set of data usable for program review and improvement processes as described in each of the case study summaries. The trajectory is apparent that as a campus embraces the use of data for intelligent decision-making, it becomes part of the culture and enables primary focus on true issues of program improvement.

As seen in the survey findings and case studies, some institutions are embracing the use of data and evaluation efforts for their graduate education community. Other media reports describe more public assessment efforts as well. At the University of Minnesota, for example, a Carnegie Foundation grant has enabled a restructured and internal approach to program review (Flaherty, 2012). Students and faculty are engaged in internal discussions focused on student assessment and outcomes for their graduate programs. Such efforts are intensive and require buy-in among the key stakeholders, as described above. Yet the outcomes can be enormously beneficial, as even months into the effort at Minnesota the initiative has shown the potential for policy and curricular changes to better align student goals, university and program resources, and outcomes.

One of the most central best practices and takeaways from the case study evaluations is that all constituencies have the potential to find value in the data collection and assessment activities. The lighthearted jab repeated by several interviewees is that everyone needs to be number one, or perhaps live in a Lake

Wobegon-type of culture where every program is above average. Going to that extreme is neither necessary nor possible. But a key point is that each program must have a hook, a reason to dig further into the data, and some desire to pay attention to the study at hand. If programs find encouragement through understanding and identifying their own strengths – and the broad range of metrics and data as described implies they can – then they can ultimately perceive value in the exercise and use it to improve themselves. This outcome lays the foundation for having broad engagement among the faculty and programs for involvement in the studies, including finding champions for these efforts. If the study outcomes are not logical and do not provide room for identifying both areas of strength and areas of weakness, then the programs will not pay attention. In short, as programs learn of their own strengths, they will enable further use for future improvements through an impetus to change.

Policy Implications

Many of these best practices and recommendations have public policy implications. Sustainable changes and quality improvements are highly difficult for institutions to cultivate, even with access to information instruments such as the NRC study or similar review efforts on campuses. By centralizing a coordinated data sharing effort through institutional priorities and leadership – policy tools and instruments in and of themselves – the onus is not solely on the programs or universities alone to generate change in isolation. Even with institutional support and leadership priorities, many of the data efforts described at the case study sites and referenced in the survey responses would not be as likely or possible without

adequate and accurate peer data. Shared efforts among the graduate education community to identify valuable metrics, common definitions, and comparative data would contribute to the institutionalization of data collection and analysis, hopefully leading to quality improvements in graduate education over time.

Assuming the national infrastructure may eventually be in place for data use and assessment efforts, through one of the entities suggested, it is still important to note that institutional structures will be critical for successful implementation at the campus level. Leadership, centralized support and resources, prioritization for these efforts, creating incentives for participation, and the identification of tangible outcomes are necessary elements for enabling the projects and studies addressed in this dissertation. Ultimately there needs to be campus engagement and buy-in to the change and improvement processes, including finding champions for change efforts on campuses and within individual graduate programs. The institutional structures leading to change need to be in place for the right reasons and not solely a push for institutional or program prestige.

Two of the three case study sites were active users of privately available data on faculty and scholarly productivity, such as those provided by third-party companies. Several survey respondent campuses alluded to or directly mentioned such data in their open-ended comments as well. The market exists for such data. Some campuses and programs are eager to review their standing on these metrics and could not obtain this information about themselves without substantial resources, much less comparative data. The benefits, however, must be tangible and easily accessible. The financial outlays required to obtain these data may move

the data out of reach for some types of institutions, especially the ones growing their faculties and research portfolios who are not as well-established as the top tier research universities.

Also, the for-profit nature of these emerging companies raises potential questions about their independence. The university clients paying for the data about themselves will need assurances that the data are accurate, yet they also need to feel the services provided are useful and worthwhile. A strong showing in the private company datasets may entice the campuses to becoming long-standing clients, thus satisfying the company's needs for continued profits. This cycle of interrelated dependencies has the potential to create dysfunctional responses for the data gathering and use. It should be noted that such outcomes may be wholly unintended with no distortion or abuse intended among the firms. It is a caution to guard against, however.

Thus coordination between the campuses, the independent conveners, and the third party providers may be necessary to achieve a truly centralized and common data set that is not wholly government-supported. Because much of the national data in a common set of metrics will be student-oriented, that portion of the data could be the key focus for conveners working to establish the framework for the common data set. The for-profit, market data available today can continue to increase in scope related to faculty productivity, research, and scholarship. The two types of data are complementary and necessary for creating an accurate and holistic picture of a graduate program's quality.

Use of assessment data and rankings studies such as the NRC can be relevant to all types of campuses and graduate programs. By reframing how these studies are viewed, the community can shift away from a focus on rankings toward one of improvement. Because of the lag in reputational factors and historical inertia related to substantive change, immediate reactions to change may not be realized. Thus these additional incentives to consider data use as campus and program improvement efforts have the potential to facilitate use plans. Longer-term goals are needed, and with the trend data that will hopefully be available, may be feasible and desired among campuses.

The role of government in such studies also has policy implications. The federal government played a large role in funding and organizing the NRC studies, driven in no small part by the desire to remain globally competitive for top talent with high-quality educational systems. Several of the recommendations and best practices described here may in fact remove that oversight at the very time that governments are asking the higher education community to be more accountable to the public, fiscally responsible, and good stewards of data. Campuses feel those pressures. Even though the federal government may no longer take a lead role in coordinating such studies – assuming it shifts to CGS, NORC, or AIR as recommended – campuses must still feel accountable.

In a manner, the recommendations and changes described here have potential for ultimately making it simpler for campuses to report accreditation and assessment data to policymakers and the general public. Sensitivity to the burden of collecting the data, validating it, and reporting it to external entities, especially in

the initial stages of this effort, should remain present. But by making it useful to campuses, programs, faculty, and students, it can ultimately be best used for policy changes and decisions.

Concluding Thoughts

This study served as a multi-level evolutional analysis with an emphasis on the institutional forces that can affect change, including external pressures such as quality and assessment studies such as the NRC. The literature and theoretical perspectives reviewed in Chapter 2 show that substantive and sustainable change is difficult for institutions of higher education. Yet, the outcomes of the surveys and case studies also show how – with adequate and committed resources, institutional priorities, and leadership – institutions could use quality data and assessments from studies such as the NRC for change and continuous improvement on campuses. Effective change processes using quality and assessment data will be motivated by institutional priorities and engagement, supported by campus champions and resources, and influenced by potential broader changes in the field, including peer comparisons on key graduate education metrics.

If the policy recommendations, data recommendations, and methodological changes for quality and assessment studies at the national level described above were to be implemented, some of the roadblocks in place today may be mitigated, enabling more campuses to develop an infrastructure of resources and priorities on data and assessment efforts. Additionally, with proper study, quality data and peer comparisons can assist in determining which factors might be high impact areas where change could most contribute to program improvement. Included in these

discussions are the private sector implications for large-scale data delivery in higher education, such as those provided by third party companies in use at increasing numbers of universities.

The recommendations to reframe quality studies have the potential to transform use away from an emphasis on inefficient and suspect rankings toward sustainable and continuous improvement. The NRC study took steps in this direction, albeit on a limited basis because of the methodology issues described in the above chapters. The survey and case study findings from this dissertation research showed how certain pockets of use – for advocacy, program review and assessment practices, student and faculty recruitment, and budgetary discussions – benefitted from the outcomes of the NRC study. Yet the findings also confirm that continuous improvement practices require focused attention, strong leadership, adequate resources, and faculty engagement.

Institutions will not necessarily make changes to improve the quality of their academic programs or overall institutional climate if they do not feel pressure from some corners to do so, whether from an inherent drive to improve and succeed or from external entities such as government oversight or accreditation and review activities. Policy and assessment tools in use today can facilitate that pressure and alleviate some of the inertia common in complex higher education organizations. Without reform and an intent to improve the rankings and assessment data and practices, the top programs will continue to be thought of as the best programs. The weaker programs may be making substantial changes to their program quality, but if the right assessment tools are not in place to reflect these changes, no one may

know, especially the prospective students and faculty who may be able to enhance program reputations long-term.

The pressures on American graduate education are increasing and the community is showing signs of coming together to ensure leadership in competitiveness and preparing students for future professions (Council of Graduate Schools, 2007). As long as traditional sources of academic funding and research become more limited, institutions will seek out ways to attract funding from other avenues – many necessitating strong showings in national rankings. Many of the best practices and recommendations made in this work can enable shifts in the outcomes of such studies. Globally, rankings continue yet each new methodology, in whichever country, encounters many of the same issues that plagued the NRC study. As oversight, cultural differences, structural and budgetary support, and data use evolve and coalesce in various countries, university decision-making and responses to rankings and assessment studies takes on greater importance.

This research suggests additional areas for future research and analysis. From a data and research perspective, longitudinal work could be important for future studies, particularly if some of the data collection efforts described surrounding graduate education data are put into place nationally. The critical need to have common definitions and indicators for graduate education across universities will be a factor in any additional research. More work can be done with the NRC study database itself, including additional analysis on the specific quality indicators. The data are minable, though the hesitations expressed in the survey results need to

be taken into consideration to ensure there is community engagement with any research findings.

From an institutional perspective, several studies conclude by calling for more work reviewing connections between institutional change and organizational practices and context (Lounsbury, 2001; Pfeffer and Salancik, 1978; Ellstrom, 1983). These topics, while difficult to capture, continue to take on increasing importance in more recent assessment literature. The surveys and case studies described in this dissertation were baseline attempts to review institutional structures that might contribute to longer-term use of rankings and assessment data. Further study is warranted.

While it is not currently known whether the NRC study will be repeated, other data collection and assessment efforts on campuses will be ongoing. Universities will always be seeking to improve themselves in various ways: their academic programs, faculty, students, and accountability to the public. Ultimately, the question to address is what can be done via public policies and assessment best practices to improve both the studies and the resulting impacts they could have on institutional quality. This dissertation hopefully provides insight into policy tools, institutional structures, and processes to contribute to long-standing improvement in doctoral education in the United States.

APPENDIX 1.1 – CHANGES TO THE NRC STUDY EXCEL DATA TABLE

Changes to the Excel Data Table for the NRC Data-Based Assessment of Research-Doctorate Programs

April 21, 2011

A revised Excel Data Table for the NRC *Data-Based Assessment of Research-Doctorate Programs in the United States* is now available. A summary of changes for each program can be found here. Those who wish to compare the September 28, 2010 version of the Data Table to the revised rankings, may find the old rankings on our website at (http://sites.nationalacademies.org/pga/Resdoc/index.htm) under the Project Information tab.

The revisions are in response to communications and queries received by the NRC since the first Data Table was released on September 28, 2010. At that time, the NRC agreed to follow up on queries about the data and these were received from approximately 450 doctoral programs from 34 institutions. Ten of these institutions had queries for 10 or more of their programs.

The most common questions centered around faculty lists and related characteristics: publications per allocated faculty member, citations per publication, the allocation of faculty, and the measure of interdisciplinarity that used this measure. The NRC was not able to permit changes in faculty lists from what universities had originally submitted. That would have required enormous expense to completely redo the study with the 2005/6 data.

In the course of this process, the NRC discovered four substantive errors. These have been corrected and incorporated into re-calculated rankings. The variables that were affected are:

1) **Average Citations per Publication.** Publications for 2002 used to obtain citations per publication had been mislabeled in all non-humanities fields. 2002 publications were corrected, and the "citations per publication" variable (which is averaged over the years 2000 to 2006) was re-calculated.

2) **Awards per Allocated Faculty Member.** The NRC undercounted honors and awards. Data for this variable were re-compiled from faculty lists and the variable was re-calculated.

3) **Percent with Academic Plans.** The response rate to this question, which was calculated from the NSF Survey of Earned Doctorates, varied considerably across

programs. It was agreed that a more accurate measure based on survey data was percent of respondents with academic positions or post-docs, not percent of total Ph.D.s. This variable was re-calculated with the changed definition.

4) **Percent of First-Year Students with Full Financial Support.** This variable had been given the value “0” when a program had no first year students. We now use an asterisk to indicate that a program has no first year students. When no data were reported, there is an "N/D".

APPENDIX 3.1 – DISSERTATION SURVEYS

Introductory Webpage

My name is Stephanie Schmitt, and I am a doctoral candidate in the Department of Public Policy at the University of North Carolina at Chapel Hill. I am conducting a research study, and this survey is part of a doctoral dissertation project on the use and benefits of national assessment projects with particular emphasis on graduate education assessment studies, such as the National Research Council's Data-Based Assessment of Research Doctorate Programs (NRC). More information about the NRC can be found at the National Academies' website.

The NRC data were collected in 2005-06, and the results were released in September 2010. The term *results* is used throughout the survey to include both the actual database and the illustrative ranges of rankings. The NRC study has admittedly been controversial; the focus of my dissertation is *not* to evaluate the NRC specifically, but rather how these types of results are used in university decision-making.

Your participation is voluntary, and the information you provide will be kept confidential. Individual responses will not be identified, and all data will be reported in aggregate. You may decline to answer any question for any reason. You may begin the survey, save it, and return to complete it at a later time, although the survey itself should take no longer than three minutes to complete.

Even if you were not in your present role during the NRC study data collection period, your responses will be appreciated. Please select responses that reflect how your university is approaching these matters.

If you have any questions about the research project or the survey itself, please contact Stephanie Schmitt at sschmitt@email.unc.edu. My advisor is Professor Maryann Feldman, feldmanm@email.unc.edu. If you have any questions about your rights as a research participant, you may contact the University of North Carolina Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu and mention study number 11-1442.

Please select the "next" arrow key below to begin this brief survey.

Central Administration Survey

Note: I do not need to ask for university affiliation because Qualtrics can capture this for me when I send out the survey invitations from within their site based on an Excel panel (with name, email, university) I upload.

1. One of the NRC's stated goals was to "provide an unparalleled dataset that can be used to assess the quality and effectiveness of doctoral programs based on measures important to faculty, students, administrators, funders, and other stakeholders". Do you believe the NRC study achieved this goal?
   ___ Yes
   ___ No

2. Thinking about the NRC study broadly on your campus, are there plans to incorporate the results?
   ___ The results have already been incorporated in campus activities and/or discussions.
   ___ There are future plans to incorporate the results in campus activities and/or discussions.
   ___ It is unknown when or if the results will be incorporated in campus activities and/or discussions.
   ___ My campus has elected not to use the results for any purpose.

(Can select more than one.)

2a. (Shows only for people who picked 2.1 or 2.2.) In which specific areas have or will the results be used broadly on your campus?

___ General conversations about key topics in graduate education within the campus
___ Campuswide benchmarking and/or strategic planning efforts
___ Program review
___ New program development priorities
___ Identifying focus areas for future data collection
___ Identifying focus areas for future data analysis
___ Accreditation and/or assessment activities
___ Budget and/or resource allocations
___ Identifying low-performing programs for further review
___ Campuswide policy decisions (e.g., research emphasis, faculty designations, student support, faculty hiring, admissions priorities)
___ Specific retreats to discuss graduate education quality and/or future directions
___ Advocating to the federal government
___ Advocating to state governments
___ Public relations purposes
___ Other: _____

(Can select more than one.)

3. One common theme at the *NRC Convocation on Analytic Uses* in March was that universities possibly found value in preparing, discussing, and collecting data for the study. In which of the following areas did your campus use the NRC study as part of the data collection process or in advance of the results being released?
   ___ *Same as list above*
   ___ None

(Can select more than one.)

4. Consider the NRC database and the NRC illustrative ranges of rankings as outcomes of the study. Both are admittedly complex and require time to analyze, understand, explain, and use them. Please respond to the following statements for your campus.

	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful.				
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful.				
We found the actual NRC database/spreadsheet useful.				
We found the NRC illustrative ranges of rankings useful.				

5. Please rate your agreement with the following statements about the NRC, including all facets of the study from data collection and assessment through the release of the results and campus discussions.

	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus had active involvement among faculty and staff for the NRC study.				
My campus had one or more champions for the NRC study.				
My campus understood the NRC study methodology and results.				
My campus was persuaded to implement change upon seeing the NRC study results.				
My campus has used or will use the NRC study results to justify decisions.				
The NRC study broadly has or will improve the quality of graduate education at my campus.				

6. If you have any additional comments about the NRC study and its use on your campus, please share them here: _____

7. In case clarifying questions arise, we may need to contact you. Our records show the following is your name and email address.

   (Qualtrics displays name and email.)

   Is that correct? Individual responses will not be identified without your prior permission.
   ___ Yes
   ___ No

8. (Shows only for people who picked 7.2/No.) If not, please provide your correct name and email address. Individual responses will not be identified without your prior permission.

First Name _____
Last Name _____
Email Address _____

9. Your campus may have an interesting story to share with others in the graduate education community. Would you be willing to be contacted for additional questions to serve as a case study for my project? Specific program names and universities will be masked in the final dissertation if requested.
   ___ Yes
   ___ No

(Must select one.)

Thank you for completing the survey and assisting a graduate student in their doctoral research!

Program-level Survey

Note: I do not need to ask for university affiliation because Qualtrics can capture this for me when I send out the survey invitations from within their site based on an Excel panel (with name, email, university) I upload.

1. One of the NRC's stated goals was to "provide an unparalleled dataset that can be used to assess the quality and effectiveness of doctoral programs based on measures important to faculty, students, administrators, funders, and other stakeholders". Do you believe the NRC study achieved this goal?
   ___ Yes
   ___ No

2. Thinking about the NRC study within the context of your specific doctoral program, are there plans to incorporate the results?
   ___ The results have already been incorporated in departmental activities and/or discussions.
   ___ There are future plans to incorporate the results in departmental activities and/or discussions.
   ___ It is unknown when or if the results will be incorporated in departmental activities and/or discussions.
   ___ My department has elected not to use the results for any purpose.

2a. (Shows only for people who picked 2.1 or 2.2.) In which specific areas have or will the results be used in your doctoral program?

___ Program review
___ Identifying focus areas for future data collection
___ Identifying focus areas for future data analysis
___ Accreditation and/or assessment activities
___ Specific retreats to discuss graduate education quality and/or future directions
___ Student recruitment and/or admissions
___ Academic/curriculum revisions (e.g., student progression, curriculum requirements)
___ Budget and resource requests to deans and/or administrators
___ Peer comparisons to identify your program's strengths and weaknesses
___ Faculty hiring plans
___ Faculty recruitment
___ Doctoral program policy revisions (e.g., funding decisions, mentoring, research activity)
___ General conversations about key topics in graduate education within the campus
___ Other: _____

(Can select more than one.)

3. One common theme at the *NRC Convocation on Analytic Uses* in March was that universities possibly found value in preparing, discussing, and collecting data for the study. In which of the following areas did your doctoral program use the NRC study as part of the data collection process or *in advance of* the results being released?
   ___ *Same as list above*
   ___ None

(Can select more than one.)

4. Consider the NRC database and the NRC illustrative ranges of rankings as outcomes of the study. Both are admittedly complex and require time to analyze, understand, explain and use them. Please respond to the following statements for your doctoral program.

	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-06 useful.				
We found activities from 2006-10 using the data and/or in preparation for the NRC release useful.				
We found the actual NRC database/spreadsheet useful.				
We found the NRC illustrative ranges of rankings useful.				

5. Consider the different types of data available in the NRC data spreadsheet. Please respond to the following statements for your doctoral program.

	Strongly Agree	Agree	Disagree	Strongly Disagree
We found the faculty productivity measures (e.g., publications, citations, awards) useful.				
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful.				
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful.				
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful.				

We found the faculty counts and allocations useful.				
We found the listings of 18 student activities and campus resources useful.				

If you have any comments about the usefulness of the various domains to place your responses above into context, please share them here: _____

6. Please rate your agreement with the following statements about the NRC, including all facets of the study from data collection and assessment through the release of the results and program discussions.

	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study.				
My graduate program had one or more champions for the NRC study.				
My graduate program understood the NRC study methodology and results.				
My graduate program was persuaded to implement change upon seeing the NRC study results.				
My graduate program has used or will use the NRC study results to justify decisions.				
The NRC study broadly has or will improve the quality my graduate program.				

7. Please rate your agreement with the following statements about the NRC study. These statements are similar to the program-level question above but are intended to gauge your thoughts on broader campus decisions regarding the NRC study.

	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results.				
My campus has used or will use the NRC results to justify decisions.				
The NRC study broadly has or will improve the quality of graduate education at my campus.				

8. If you have any additional comments about the NRC study and its use on your campus, please share them here: _____

9. In case clarifying questions arise, we may need to contact you. Our records show the following is your name and email address.

   (Qualtrics displays name and email.)

   Is that correct? Individual responses will not be identified without your prior permission.
   ___ Yes
   ___ No

10. (Shows only for people who picked 9.2/No.) If not, please provide your correct name and email address. Individual responses will not be identified without your prior permission.

First Name _____
Last Name _____
Email Address _____

11. Your program may have an interesting story to share with others in the graduate education community. Would you be willing to be contacted for additional questions to serve as a case study for my project? Specific program names and universities will be masked in the final dissertation if requested.
    ___ Yes
    ___ No

(Must select one.)

Thank you for completing the survey and assisting a graduate student in their doctoral research!

APPENDIX 3.2 – DISSERTATION SURVEY COMMUNICATIONS

All Email communications can be sent out via the Qualtrics survey software. They will come from my name and email address and will be individual emails, i.e., the respondents do not see a mass email message with all names.

Email 1: send out from Qualtrics one week prior to the emailed survey invitation

Subject: Share your Views about Graduate Education

Dr. ${m://FirstName} ${m://LastName}
${e://Field/Title}
${e://Field/University}

Dear Dr. ${m://LastName}:

My name is Stephanie Schmitt, and I am writing to ask you to participate in a short online survey of {Provosts and Chief Academic Officers} {Chairs (and Program Directors) in the X field} for my doctoral dissertation project. My dissertation will review the use and benefits of national assessment/rankings projects with particular emphasis on graduate education studies, such as the National Research Council's Data-Based Assessment of Research Doctorate Programs (NRC).

The survey should take no longer than three {or four} minutes to complete. In the next week, the invitation to the electronic survey will come from my email address via Qualtrics with a subject line of "Invitation to Share Opinions on Graduate Education".

Your responses will be invaluable to my research and are greatly appreciated. My hope is that the results of my dissertation will contribute to discussions on the quality of graduate education in the United States.

I will share a short report with respondents that provides an overview of the survey results from universities around the country. Thank you in advance for helping a doctoral student on their dissertation project.

Best regards,
Stephanie Schmitt
UNC-Chapel Hill, Doctoral Candidate in Public Policy

Email 2: survey invitation from Qualtrics

Subject: Invitation to Share Opinions on Graduate Education

Dear Dr. ${m://LastName}:

Last week, I sent a note inviting your participation in a three {or four}-minute online survey as part of my dissertation project on graduate education and the NRC assessment of doctoral programs.

Please click the individualized link below to begin the survey:

${l://SurveyLink?d=Take the Survey}

My dissertation project will assess the use and benefits of national assessment/rankings projects with particular emphasis on graduate education studies, such as the National Research Council's Data-Based Assessment of Research Doctorate Programs (NRC).

I am conducting surveys of university {Provosts and Chief Academic Officers} {Chairs (and Program Directors) in the X field} to gather opinions on the NRC. Given the small sample nationally, it is very important I hear from you.

Your responses will be invaluable to my research and are greatly appreciated. My hope is that the results of my dissertation will contribute to discussions on the quality of graduate education in the United States.

I will share a short report with respondents that provides an overview of the survey results from universities around the country. Thank you in advance for helping a doctoral student on their dissertation project.

Best regards,
Stephanie Schmitt
UNC-Chapel Hill, Doctoral Candidate in Public Policy

If the link above did not work, please copy and paste the full survey URL below into your internet browser:
${l://SurveyURL}

Email 3: reminder from Qualtrics after one week

Subject: Reminder about Graduate Education Survey

Dear Dr. ${m://LastName}:

About a week ago I asked if you would please complete a brief survey on the use and benefits of national assessment/rankings projects such as the NRC. I have not yet received your response and would very much appreciate hearing from you.

In order for my dissertation project to represent all {Provosts and Chief Academic Officers} {Chairs (and Program Directors) in the X field}, I really need your participation. I hope you will take three {or four} short minutes to click the link below and complete the survey.

${l://SurveyLink?d=Take the Survey}

Your responses will be invaluable to my research and are greatly appreciated. My hope is that the results of my dissertation will contribute to discussions on the quality of graduate education in the United States.

I will share a short report with respondents that provides an overview of the survey results from universities around the country. Thank you in advance for helping a doctoral student on their dissertation project.

Best regards,
Stephanie Schmitt
UNC-Chapel Hill, Doctoral Candidate in Public Policy

If the link above did not work, please copy and paste the full survey URL below into your internet browser:
${l://SurveyURL}

Email 4: final reminder from Qualtrics two weeks later

Subject: Final Chance to Provide Assistance to a Doctoral Student

Dear Dr. ${m://LastName}:

Over X% of your fellow {Provosts and Chief Academic Officers} {Chairs (and Program Directors) in the X field} have completed my online survey on the use and benefits of national assessment/rankings projects such as the NRC. Won't you please join them by clicking the link below to complete the survey?

${l://SurveyLink?d=Take the Survey}

The survey should take no longer than three {or four} minutes to complete. In order for my dissertation project to represent all {Provosts and Chief Academic Officers} {Chairs (and Program Directors) in the X field}, I really need your participation.

Your responses will be invaluable to my research and are greatly appreciated. My hope is that the results of my dissertation will contribute to discussions on the quality of graduate education in the United States.

I will share a short report with respondents that provides an overview of the survey results from universities around the country. Thank you in advance for helping a doctoral student on their dissertation project.

Best regards,
Stephanie Schmitt
UNC-Chapel Hill, Doctoral Candidate in Public Policy

If the link above did not work, please copy and paste the full survey URL below into your internet browser:
${l://SurveyURL}

APPENDIX 3.3 – DISSERTATION SURVEY TIMING

1) Central Administration Survey

Survey Phases	Actual Date
Email Communication (#1)	9/1/11 – 9:17am
Survey Invitation (#2)	9/7/11 – 10:18pm
Survey Reminder 1 (#3)	9/13/11 – 10:33pm
Survey Reminder 2 (#4)	9/27/11 – 9:56pm
Close/Inactivate Survey	12/2/11

2) Nutrition Survey

Survey Phases	Actual Date
Email Communication (#1)	9/7/11 – 9:51am
Survey Invitation (#2)	9/14/11 – 6:51am
Survey Reminder 1 (#3)	9/21/11 – 12:03pm
Survey Reminder 2 (#4)	10/5/11 – 9:56pm
Close/Inactivate Survey	12/2/11

3) Neuroscience and Neurobiology Survey

Survey Phases	Actual Date
Email Communication (#1)	9/14/11 – 9:07am
Survey Invitation (#2)	9/21/11 – 4:36am
Survey Reminder 1 (#3)	9/28/11 – 12:03pm
Survey Reminder 2 (#4)	10/12/11 – 9:56pm
Close/Inactivate Survey	12/2/11

4) Materials Science and Engineering Survey

Survey Phases	Actual Date
Email Communication (#1)	9/21/11 – 7:05pm
Survey Invitation (#2)	9/28/11 – 6:20 am
Survey Reminder 1 (#3)	10/5/11 – 12:03pm
Survey Reminder 2 (#4)	10/19/11 – 9:56pm
Close/Inactivate Survey	12/2/11

5) English Language and Literature Survey

Survey Phases	Actual Date
Email Communication (#1)	9/28/11 – 8:36 pm
Survey Invitation (#2)	10/5/11 – 6:20am
Survey Reminder 1 (#3)	10/12/11 – 12:03pm
Survey Reminder 2 (#4)	10/26/11 – 9:56pm
Close/Inactivate Survey	12/2/11

6) Chemistry Survey

Survey Phases	Actual Date
Email Communication (#1)	10/13/11 – 8:42pm
Survey Invitation (#2)	10/19/11 – 6:20am
Survey Reminder 1 (#3)	10/26/11 – 12:03pm
Survey Reminder 2 (#4)	11/9/11 – 9:56pm
Close/Inactivate Survey	12/2/11

7) Economics Survey

Survey Phases	Actual Date
Email Communication (#1)	10/18/11 – 9:20pm
Survey Invitation (#2)	10/25/11 – 6:20am
Survey Reminder 1 (#3)	11/1/11 – 12:03pm
Survey Reminder 2 (#4)	11/15/11 – 9:56pm
Close/Inactivate Survey	12/2/11

APPENDIX 4.1 – CENTRAL ADMINISTRATION SURVEY RESULTS

Table A4.1-1 – Areas of Use for Central Administration Survey Respondents

Area of Use	% Respondents	N Respondents
General conversations about key topics in graduate education within the campus	75%	39
Program review	69%	36
Campuswide benchmarking and/or strategic planning efforts	56%	29
Identifying low-performing programs for further review	39%	20
Identifying focus areas for future data collection	35%	18
Accreditation and/or assessment activities	29%	15
Budget and/or resource allocations	29%	15
Campuswide policy decisions	29%	15
Public relations purposes	27%	14
Identifying focus areas for future data analysis	25%	13
Specific retreats to discuss graduate education quality and/or future directions	15%	8
Advocating to state governments	10%	5
New program development priorities	6%	3
Advocating to the federal government	2%	1

n=52

Table A4.1-2 – Difference in Areas of Use in Advance of Results Release for Central Administration Survey Respondents

Area of Use	N Respondents	Difference
General conversations about key topics in graduate education within the campus	44	5
Program review	31	-5
Campuswide benchmarking and/or strategic planning efforts	25	-4
Identifying focus areas for future data collection	25	7
Identifying focus areas for future data analysis	22	9
Identifying low-performing programs for further review	18	-2
Campuswide policy decisions	12	-3
Public relations purposes	11	-3
Accreditation and/or assessment activities	10	-5
Specific retreats to discuss graduate education quality and/or future directions	10	2
Budget and/or resource allocations	9	-6
New program development priorities	3	-
Advocating to the federal government	1	-
Advocating to state governments	0	-5

Table A4.1-3 – Central Administration Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=99)	5% (5)	55% (54)	28% (28)	12% (12)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=100)	3% (3)	42% (42)	39% (39)	16% (16)
We found the actual NRC database/spreadsheet useful. (n=101)	5% (5)	55% (56)	31% (31)	9% (9)
We found the NRC illustrative ranges of rankings useful. (n=101)	7% (7)	39% (39)	39% (39)	16% (16)

Table A4.1-4 – Central Administration Rankings of NRC Study Use Factors on Campus

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus had active involvement among faculty and staff for the NRC study. (n=100)	15% (15)	52% (52)	27% (27)	6% (6)
My campus had one or more champions for the NRC study. (n=98)	15% (15)	61% (60)	17% (17)	6% (6)
My campus understood the NRC study methodology and results. (n=100)	12% (12)	51% (51)	31% (31)	6% (6)
My campus was persuaded to implement change upon seeing the NRC study results. (n=100)	1% (1)	16% (16)	61% (61)	22% (22)
My campus has used or will use the NRC study results to justify decisions. (n=99)	4% (4)	36% (36)	37% (37)	22% (22)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=98)	3% (3)	29% (28)	47% (46)	21% (21)

APPENDIX 4.2 – COMBINED PROGRAMS SURVEY RESULTS

Table A4.2-1 – Areas of Use for Doctoral Program Survey Respondents

Area of Use	% Respondents	N Respondents
Program review	66%	91
Peer comparisons to identify your program's strengths and weaknesses	61%	84
Student recruitment and/or admissions	51%	70
Budget and resource requests to deans and/or administrators	45%	62
General conversations about key topics in graduate education within the campus	39%	53
Doctoral program policy revisions	30%	41
Faculty hiring plans	27%	37
Identifying focus areas for future data collection	26%	36
Specific retreats to discuss graduate education quality and/or future directions	26%	35
Faculty recruitment	24%	33
Academic/curriculum revisions	23%	32
Identifying focus areas for future data analysis	22%	30
Accreditation and/or assessment activities	18%	24

n=137

Table A4.2-2 – Difference in Areas of Use in Advance of Results Release for Doctoral Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	103	12
Peer comparisons to identify your program's strengths and weaknesses	81	-3
General conversations about key topics in graduate education within the campus	55	2
Student recruitment and/or admissions	46	-24
Budget and resource requests to deans and/or administrators	45	-17
Doctoral program policy revisions	42	1
Accreditation and/or assessment activities	38	14
Identifying focus areas for future data collection	37	1
Academic/curriculum revisions	35	3
Identifying focus areas for future data analysis	31	1
Specific retreats to discuss graduate education quality and/or future directions	25	-10
Faculty hiring plans	25	-12
Faculty recruitment	22	-11

Table A4.2-3 – Doctoral Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=312)	4% (11)	43% (135)	41% (128)	12% (38)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=309)	3% (8)	27% (83)	56% (174)	14% (44)
We found the actual NRC database/spreadsheet useful. (n=314)	5% (17)	39% (123)	38% (120)	17% (54)
We found the NRC illustrative ranges of rankings useful. (n=314)	5% (14)	37% (116)	41% (127)	18% (57)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=307)	12% (37)	48% (147)	31% (96)	9% (27)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=307)	7% (22)	51% (156)	32% (99)	10% (30)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=306)	7% (21)	49% (151)	36% (110)	8% (24)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=307)	4% (11)	48% (147)	39% (121)	9% (28)
We found the faculty counts and allocations useful. (n=305)	5% (14)	40% (121)	44% (134)	12% (36)
We found the listings of 18 student activities and campus resources useful. (n=306)	1% (3)	23% (69)	62% (190)	14% (44)

Table A4.2-4 – Doctoral Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=300)	4% (12)	38% (114)	44% (133)	14% (41)
My graduate program had one or more champions for the NRC study. (n=297)	3% (10)	20% (58)	55% (164)	22% (65)
My graduate program understood the NRC study methodology and results. (n=302)	5% (14)	38% (114)	42% (128)	15% (46)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=301)	1% (2)	14% (42)	62% (187)	23% (70)
My graduate program has used or will use the NRC study results to justify decisions. (n=302)	3% (8)	35% (107)	43% (129)	19% (58)
The NRC study broadly has or will improve the quality of my graduate program. (n=301)	3% (8)	23% (68)	52% (157)	23% (68)

Table A4.2-5 – Doctoral Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=301)	2% (5)	14% (42)	69% (207)	16% (47)
My campus has used or will use the NRC results to justify decisions. (n=297)	3% (9)	39% (116)	43% (129)	15% (43)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=300)	2% (5)	19% (58)	58% (173)	21% (64)

APPENDIX 4.3 – SURVEY RESULTS BASED ON QUALITY RANKINGS

Table A4.3-1 – Areas of Use for High-Quality Doctoral Program Survey Respondents

Area of Use	% Respondents	N Respondents
Program review	74%	48
Peer comparisons to identify your program's strengths and weaknesses	59%	38
Student recruitment and/or admissions	55%	36
Budget and resource requests to deans and/or administrators	49%	32
General conversations about key topics in graduate education within the campus	34%	22
Faculty recruitment	26%	17
Specific retreats to discuss graduate education quality and/or future directions	23%	15
Faculty hiring plans	22%	14
Identifying focus areas for future data collection	22%	14
Academic/curriculum revisions	20%	13
Doctoral program policy revisions	20%	13
Accreditation and/or assessment activities	15%	10
Identifying focus areas for future data analysis	14%	9

n=65

Table A4.3-2 – Difference in Areas of Use in Advance of Results Release for High-Quality Doctoral Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	48	0
Peer comparisons to identify your program's strengths and weaknesses	32	-6
General conversations about key topics in graduate education within the campus	24	2
Student recruitment and/or admissions	23	-13
Budget and resource requests to deans and/or administrators	19	-13
Accreditation and/or assessment activities	16	6
Academic/curriculum revisions	14	1
Doctoral program policy revisions	14	1
Identifying focus areas for future data analysis	11	2
Identifying focus areas for future data collection	11	-3
Specific retreats to discuss graduate education quality and/or future directions	11	-4
Faculty recruitment	9	-8
Faculty hiring plans	7	-7

Table A4.3-3 – High-Quality Doctoral Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=120)	6% (7)	46% (55)	39% (47)	9% (11)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=120)	5% (6)	33% (39)	51% (61)	12% (14)
We found the actual NRC database/spreadsheet useful. (n=121)	9% (11)	41% (50)	39% (47)	11% (13)
We found the NRC illustrative ranges of rankings useful. (n=120)	7% (8)	49% (59)	33% (40)	11% (13)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=119)	15% (18)	52% (62)	27% (32)	6% (7)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=119)	9% (11)	54% (64)	29% (35)	8% (9)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=119)	8% (9)	55% (65)	31% (37)	7% (8)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=119)	4% (5)	48% (57)	40% (48)	8% (9)
We found the faculty counts and allocations useful. (n=119)	7% (8)	36% (43)	48% (57)	9% (11)
We found the listings of 18 student activities and campus resources useful. (n=119)	2% (2)	23% (27)	61% (73)	14% (17)

Table A4.3-4 – High-Quality Doctoral Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=117)	6% (7)	47% (55)	40% (47)	7% (8)
My graduate program had one or more champions for the NRC study. (n=114)	4% (5)	27% (31)	54% (61)	15% (17)
My graduate program understood the NRC study methodology and results. (n=117)	7% (8)	44% (51)	36% (42)	14% (16)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=116)	2% (2)	12% (14)	64% (74)	22% (26)
My graduate program has used or will use the NRC study results to justify decisions. (n=116)	6% (7)	38% (44)	39% (45)	17% (20)
The NRC study broadly has or will improve the quality of my graduate program. (n=117)	6% (7)	24% (28)	50% (59)	20% (23)

Table A4.3-5 – High-Quality Doctoral Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=117)	3% (4)	15% (17)	68% (79)	15% (17)
My campus has used or will use the NRC results to justify decisions. (n=116)	5% (6)	41% (47)	41% (48)	13% (15)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=117)	3% (3)	23% (27)	56% (66)	18% (21)

APPENDIX 4.4 – NUTRITION SURVEY RESULTS

Table A4.4-1 – Areas of Use for Nutrition Program Survey Respondents

Area of Use	% Respondents	N Respondents
Program review	82%	9
Peer comparisons to identify your program's strengths and weaknesses	55%	6
Student recruitment and/or admissions	55%	6
Doctoral program policy revisions	45%	5
General conversations about key topics in graduate education within the campus	36%	4
Identifying focus areas for future data collection	36%	4
Academic/curriculum revisions	27%	3
Budget and resource requests to deans and/or administrators	27%	3
Identifying focus areas for future data analysis	27%	3
Specific retreats to discuss graduate education quality and/or future directions	27%	3
Accreditation and/or assessment activities	18%	2
Faculty hiring plans	18%	2
Faculty recruitment	18%	2

n=11

Table A4.4-2 – Difference in Areas of Use in Advance of Results Release for Nutrition Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	10	1
General conversations about key topics in graduate education within the campus	7	3
Doctoral program policy revisions	5	-
Student recruitment and/or admissions	5	-1
Identifying focus areas for future data collection	4	-
Peer comparisons to identify your program's strengths and weaknesses	4	-2
Accreditation and/or assessment activities	3	1
Identifying focus areas for future data analysis	3	-
Academic/curriculum revisions	2	-1
Budget and resource requests to deans and/or administrators	2	-1
Specific retreats to discuss graduate education quality and/or future directions	2	-1
Faculty hiring plans	1	-1
Faculty recruitment	1	-1

Table A4.4-3 – Nutrition Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=19)	5% (1)	37% (7)	53% (10)	5% (1)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=19)	0% (0)	26% (5)	68% (13)	5% (1)
We found the actual NRC database/spreadsheet useful. (n=19)	0% (0)	42% (8)	47% (9)	11% (2)
We found the NRC illustrative ranges of rankings useful. (n=19)	0% (0)	42% (8)	47% (9)	11% (2)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=19)	16% (3)	47% (9)	32% (6)	5% (1)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=19)	16% (3)	53% (10)	26% (5)	5% (1)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=19)	11% (2)	42% (8)	42% (8)	5% (1)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=19)	5% (1)	21% (4)	63% (12)	11% (2)
We found the faculty counts and allocations useful. (n=19)	0% (0)	26% (5)	63% (12)	11% (2)
We found the listings of 18 student activities and campus resources useful. (n=19)	0% (0)	21% (4)	68% (13)	11% (2)

Table A4.4-4 – Nutrition Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=19)	5% (1)	16% (3)	53% (10)	26% (5)
My graduate program had one or more champions for the NRC study. (n=19)	5% (1)	37% (7)	32% (6)	26% (5)
My graduate program understood the NRC study methodology and results. (n=19)	0% (0)	53% (10)	26% (5)	21% (4)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=19)	0% (0)	11% (2)	68% (13)	21% (4)
My graduate program has used or will use the NRC study results to justify decisions. (n=19)	0% (0)	42% (8)	47% (9)	11% (2)
The NRC study broadly has or will improve the quality of my graduate program. (n=19)	0% (0)	37% (7)	47% (9)	16% (3)

Table A4.4-5 – Nutrition Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=19)	0% (0)	16% (3)	63% (12)	21% (4)
My campus has used or will use the NRC results to justify decisions. (n=19)	0% (0)	21% (4)	58% (11)	21% (4)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=19)	0% (0)	32% (6)	42% (8)	26% (5)

APPENDIX 4.5 – NEUROSCIENCE AND NEUROBIOLOGY SURVEY RESULTS

Table A4.5-1 – Areas of Use for Neuroscience Program Survey Respondents

Area of Use	% Respondents	N Respondents
General conversations about key topics in graduate education within the campus	58%	11
Program review	58%	11
Student recruitment and/or admissions	47%	9
Identifying focus areas for future data analysis	42%	8
Peer comparisons to identify your program's strengths and weaknesses	42%	8
Identifying focus areas for future data collection	37%	7
Academic/curriculum revisions	21%	4
Accreditation and/or assessment activities	11%	2
Faculty recruitment	11%	2
Budget and resource requests to deans and/or administrators	5%	1
Doctoral program policy revisions	5%	1
Faculty hiring plans	5%	1
Specific retreats to discuss graduate education quality and/or future directions	5%	1

n=19

Table A4.5-2 – Difference in Areas of Use in Advance of Results Release for Neuroscience Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	17	6
Peer comparisons to identify your program's strengths and weaknesses	12	4
Identifying focus areas for future data collection	11	4
General conversations about key topics in graduate education within the campus	10	-1
Accreditation and/or assessment activities	8	6
Doctoral program policy revisions	7	6
Identifying focus areas for future data analysis	7	-1
Specific retreats to discuss graduate education quality and/or future directions	6	5
Student recruitment and/or admissions	5	-4
Academic/curriculum revisions	4	-
Budget and resource requests to deans and/or administrators	3	2
Faculty hiring plans	1	-
Faculty recruitment	1	-1

Table A4.5-3 – Neuroscience Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=54)	0% (0)	44% (24)	39% (21)	17% (9)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=54)	2% (1)	31% (17)	52% (28)	15% (8)
We found the actual NRC database/spreadsheet useful. (n=54)	0% (0)	30% (16)	46% (25)	24% (13)
We found the NRC illustrative ranges of rankings useful. (n=54)	2% (1)	31% (17)	50% (27)	17% (9)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=50)	6% (3)	54% (27)	22% (11)	18% (9)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=50)	10% (5)	52% (26)	28% (14)	10% (5)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=50)	8% (4)	50% (25)	30% (15)	12% (6)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=50)	2% (1)	46% (23)	32% (16)	20% (10)
We found the faculty counts and allocations useful. (n=50)	0% (0)	26% (13)	54% (27)	20% (10)
We found the listings of 18 student activities and campus resources useful. (n=50)	0% (0)	28% (14)	58% (29)	14% (7)

Table A4.5-4 – Neuroscience Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=50)	4% (2)	38% (19)	42% (21)	16% (8)
My graduate program had one or more champions for the NRC study. (n=49)	4% (2)	31% (15)	39% (19)	26% (13)
My graduate program understood the NRC study methodology and results. (n=50)	2% (1)	30% (15)	50% (25)	18% (9)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=50)	0% (0)	12% (6)	54% (27)	34% (17)
My graduate program has used or will use the NRC study results to justify decisions. (n=50)	0% (0)	24% (12)	44% (22)	32% (16)
The NRC study broadly has or will improve the quality of my graduate program. (n=49)	0% (0)	14% (7)	57% (28)	29% (14)

Table A4.5-5 – Neuroscience Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=50)	2% (1)	16% (8)	66% (33)	16% (8)
My campus has used or will use the NRC results to justify decisions. (n=50)	4% (2)	34% (17)	48% (24)	14% (7)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=50)	0% (0)	8% (4)	64% (32)	28% (14)

APPENDIX 4.6 – MATERIALS SCIENCE SURVEY RESULTS

Table A4.6-1 – Areas of Use for Materials Science Program Survey Respondents

Area of Use	% Respondents	N Respondents
Peer comparisons to identify your program's strengths and weaknesses	91%	10
Specific retreats to discuss graduate education quality and/or future directions	73%	8
Program review	64%	7
Faculty hiring plans	55%	6
Faculty recruitment	55%	6
Student recruitment and/or admissions	55%	6
Budget and resource requests to deans and/or administrators	45%	5
General conversations about key topics in graduate education within the campus	45%	5
Academic/curriculum revisions	36%	4
Identifying focus areas for future data analysis	36%	4
Identifying focus areas for future data collection	36%	4
Doctoral program policy revisions	27%	3
Accreditation and/or assessment activities	18%	2

n=11

Table A4.6-2 – Difference in Areas of Use in Advance of Results Release for Materials Science Program Survey Respondents

Area of Use	N Respondents	Difference
Peer comparisons to identify your program's strengths and weaknesses	12	2
Program review	11	4
Academic/curriculum revisions	9	5
Specific retreats to discuss graduate education quality and/or future directions	9	1
Student recruitment and/or admissions	8	2
Faculty hiring plans	7	1
Faculty recruitment	7	1
Accreditation and/or assessment activities	6	4
Budget and resource requests to deans and/or administrators	6	1
General conversations about key topics in graduate education within the campus	6	1
Doctoral program policy revisions	5	2
Identifying focus areas for future data analysis	4	-
Identifying focus areas for future data collection	4	-

Table A4.6-3 – Materials Science Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=32)	6% (2)	34% (11)	41% (13)	19% (6)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=32)	6% (2)	22% (7)	47% (15)	25% (8)
We found the actual NRC database/spreadsheet useful. (n=32)	9% (3)	38% (12)	38% (12)	16% (5)
We found the NRC illustrative ranges of rankings useful. (n=33)	9% (3)	33% (11)	42% (14)	15% (5)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=33)	21% (7)	46% (15)	24% (8)	9% (3)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=33)	6% (2)	55% (18)	27% (9)	12% (4)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=32)	3% (1)	47% (15)	41% (13)	9% (3)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=33)	3% (1)	46% (15)	39% (13)	12% (4)
We found the faculty counts and allocations useful. (n=31)	6% (2)	29% (9)	52% (16)	13% (4)
We found the listings of 18 student activities and campus resources useful. (n=32)	3% (1)	31% (10)	53% (17)	13% (4)

Table A4.6-4 – Materials Science Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=32)	3% (1)	41% (13)	34% (11)	22% (7)
My graduate program had one or more champions for the NRC study. (n=32)	0% (0)	13% (4)	59% (19)	28% (9)
My graduate program understood the NRC study methodology and results. (n=32)	3% (1)	31% (10)	41% (13)	25% (8)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=30)	0% (0)	23% (7)	43% (13)	33% (10)
My graduate program has used or will use the NRC study results to justify decisions. (n=32)	0% (0)	31% (10)	47% (15)	22% (7)
The NRC study broadly has or will improve the quality of my graduate program. (n=31)	0% (0)	29% (9)	39% (12)	32% (10)

Table A4.6-5 – Materials Science Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=32)	3% (1)	6% (2)	69% (22)	22% (7)
My campus has used or will use the NRC results to justify decisions. (n=32)	3% (1)	31% (10)	47% (15)	19% (6)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=32)	6% (2)	19% (6)	47% (15)	28% (9)

APPENDIX 4.7 – ENGLISH LANGUAGE AND LITERATURE SURVEY RESULTS

Table A4.7-1 – Areas of Use for English Program Survey Respondents

Area of Use	% Respondents	N Respondents
Student recruitment and/or admissions	54%	15
Budget and resource requests to deans and/or administrators	50%	14
Peer comparisons to identify your program's strengths and weaknesses	46%	13
Program review	46%	13
General conversations about key topics in graduate education within the campus	36%	10
Doctoral program policy revisions	29%	8
Identifying focus areas for future data collection	29%	8
Faculty hiring plans	25%	7
Academic/curriculum revisions	18%	5
Accreditation and/or assessment activities	18%	5
Faculty recruitment	14%	4
Identifying focus areas for future data analysis	14%	4
Specific retreats to discuss graduate education quality and/or future directions	7%	2

n=28

Table A4.7-2 – Difference in Areas of Use in Advance of Results Release for English Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	18	5
General conversations about key topics in graduate education within the campus	14	4
Peer comparisons to identify your program's strengths and weaknesses	13	-
Identifying focus areas for future data collection	10	2
Student recruitment and/or admissions	10	-5
Doctoral program policy revisions	8	-
Identifying focus areas for future data analysis	7	3
Academic/curriculum revisions	6	1
Accreditation and/or assessment activities	6	1
Budget and resource requests to deans and/or administrators	5	-9
Faculty recruitment	4	-
Faculty hiring plans	3	-4
Specific retreats to discuss graduate education quality and/or future directions	2	-

Table A4.7-3 – English Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=57)	7% (4)	39% (22)	40% (23)	14% (8)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=58)	3% (2)	28% (16)	50% (29)	19% (11)
We found the actual NRC database/spreadsheet useful. (n=58)	9% (5)	29% (17)	36% (21)	26% (15)
We found the NRC illustrative ranges of rankings useful. (n=57)	9% (5)	30% (17)	32% (18)	30% (17)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=57)	9% (5)	37% (21)	46% (26)	9% (5)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=57)	7% (4)	54% (31)	28% (16)	11% (6)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=57)	9% (5)	54% (31)	30% (17)	7% (4)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=57)	4% (2)	49% (28)	40% (23)	7% (4)
We found the faculty counts and allocations useful. (n=57)	5% (3)	42% (24)	42% (24)	11% (6)
We found the listings of 18 student activities and campus resources useful. (n=57)	0% (0)	23% (13)	63% (36)	14% (8)

Table A4.7-4 – English Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=57)	9% (5)	49% (28)	33% (19)	9% (5)
My graduate program had one or more champions for the NRC study. (n=54)	7% (4)	19% (10)	52% (28)	22% (12)
My graduate program understood the NRC study methodology and results. (n=57)	5% (3)	42% (24)	39% (22)	14% (8)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=57)	0% (0)	14% (8)	60% (34)	26% (15)
My graduate program has used or will use the NRC study results to justify decisions. (n=57)	7% (4)	28% (16)	39% (22)	26% (15)
The NRC study broadly has or will improve the quality of my graduate program. (n=57)	9% (5)	18% (10)	42% (24)	32% (18)

Table A4.7-5 – English Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=57)	0% (0)	18% (10)	63% (36)	19% (11)
My campus has used or will use the NRC results to justify decisions. (n=57)	0% (0)	42% (24)	37% (21)	21% (12)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=57)	2% (1)	23% (13)	53% (30)	23% (13)

APPENDIX 4.8 – CHEMISTRY SURVEY RESULTS

Table A4.8-1 – Areas of Use for Chemistry Program Survey Respondents

Area of Use	% Respondents	N Respondents
Peer comparisons to identify your program's strengths and weaknesses	80%	36
Program review	78%	35
Budget and resource requests to deans and/or administrators	56%	25
Student recruitment and/or admissions	53%	24
Doctoral program policy revisions	36%	16
Faculty hiring plans	36%	16
Specific retreats to discuss graduate education quality and/or future directions	33%	15
General conversations about key topics in graduate education within the campus	29%	13
Academic/curriculum revisions	27%	12
Faculty recruitment	27%	12
Identifying focus areas for future data collection	22%	10
Accreditation and/or assessment activities	16%	7
Identifying focus areas for future data analysis	16%	7

n=45

Table A4.8-2 – Difference in Areas of Use in Advance of Results Release for Chemistry Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	34	-34
Peer comparisons to identify your program's strengths and weaknesses	28	-8
Budget and resource requests to deans and/or administrators	18	-7
General conversations about key topics in graduate education within the campus	13	-
Student recruitment and/or admissions	12	-12
Academic/curriculum revisions	10	-2
Accreditation and/or assessment activities	10	3
Doctoral program policy revisions	10	-6
Faculty hiring plans	10	-6
Identifying focus areas for future data analysis	8	1
Identifying focus areas for future data collection	7	-3
Faculty recruitment	6	-6
Specific retreats to discuss graduate education quality and/or future directions	5	-10

Table A4.8-3 – Chemistry Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=91)	1% (1)	56% (51)	34% (31)	9% (8)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=88)	1% (1)	33% (29)	55% (48)	11% (10)
We found the actual NRC database/spreadsheet useful. (n=91)	5% (5)	49% (45)	31% (28)	14% (13)
We found the NRC illustrative ranges of rankings useful. (n=91)	4% (4)	40% (36)	40% (36)	16% (15)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=87)	15% (13)	57% (50)	22% (19)	6% (5)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=87)	7% (6)	52% (45)	31% (27)	10% (9)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=87)	3% (3)	53% (46)	34% (30)	9% (8)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=87)	3% (3)	52% (45)	38% (33)	7% (6)
We found the faculty counts and allocations useful. (n=87)	8% (7)	51% (44)	32% (28)	9% (8)
We found the listings of 18 student activities and campus resources useful. (n=87)	1% (1)	23% (20)	58% (50)	18% (16)

Table A4.8-4 – Chemistry Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=84)	2% (2)	42% (35)	49% (41)	7% (6)
My graduate program had one or more champions for the NRC study. (n=85)	1% (1)	19% (16)	65% (55)	15% (13)
My graduate program understood the NRC study methodology and results. (n=87)	2% (2)	39% (34)	43% (37)	16% (14)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=87)	1% (1)	13% (11)	72% (63)	14% (12)
My graduate program has used or will use the NRC study results to justify decisions. (n=87)	1% (1)	52% (45)	33% (29)	14% (12)
The NRC study broadly has or will improve the quality of my graduate program. (n=87)	1% (1)	23% (20)	61% (53)	15% (13)

Table A4.8-5 – Chemistry Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=86)	2% (2)	14% (12)	73% (63)	11% (9)
My campus has used or will use the NRC results to justify decisions. (n=85)	5% (4)	45% (38)	40% (34)	11% (9)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=85)	1% (1)	21% (18)	62% (53)	15% (13)

APPENDIX 4.9 – ECONOMICS SURVEY RESULTS

Table A4.9-1 – Areas of Use for Economics Program Survey Respondents

Area of Use	% Respondents	N Respondents
Program review	70%	16
Budget and resource requests to deans and/or administrators	61%	14
Peer comparisons to identify your program's strengths and weaknesses	48%	11
General conversations about key topics in graduate education within the campus	44%	10
Student recruitment and/or admissions	44%	10
Doctoral program policy revisions	35%	8
Faculty recruitment	30%	7
Accreditation and/or assessment activities	26%	6
Specific retreats to discuss graduate education quality and/or future directions	26%	6
Faculty hiring plans	22%	5
Academic/curriculum revisions	17%	4
Identifying focus areas for future data analysis	17%	4
Identifying focus areas for future data collection	13%	3

n=23

Table A4.9-2 – Difference in Areas of Use in Advance of Results Release for Economics Program Survey Respondents

Area of Use	N Respondents	Difference
Program review	13	-3
Peer comparisons to identify your program's strengths and weaknesses	12	1
Budget and resource requests to deans and/or administrators	11	-3
Doctoral program policy revisions	7	-1
Student recruitment and/or admissions	6	-4
Accreditation and/or assessment activities	5	-1
General conversations about key topics in graduate education within the campus	5	-5
Academic/curriculum revisions	4	-
Faculty hiring plans	3	-2
Faculty recruitment	3	-4
Identifying focus areas for future data analysis	2	-2
Identifying focus areas for future data collection	1	-2
Specific retreats to discuss graduate education quality and/or future directions	1	-5

Table A4.9-3 – Economics Program Rankings of Usefulness of NRC Study Elements

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
We found collecting and submitting the campus data for the NRC in 2005-2006 useful. (n=59)	5% (3)	34% (20)	51% (30)	10% (6)
We found activities from 2006-2010 using the data and/or in preparation for the NRC release useful. (n=58)	3% (2)	16% (9)	71% (41)	10% (6)
We found the actual NRC database/spreadsheet useful. (n=60)	7% (4)	42% (25)	42% (25)	10% (6)
We found the NRC illustrative ranges of rankings useful. (n=60)	2% (1)	45% (27)	38% (23)	15% (9)
We found the faculty productivity measures (e.g., publications, citations, awards) useful. (n=61)	10% (6)	41% (25)	43% (26)	7% (4)
We found the diversity measures (e.g., minority and female faculty, minority, female and international students) useful. (n=61)	3% (2)	43% (26)	46% (28)	8% (5)
We found the student support and outcomes measures (e.g., financial support, completion rates, time to degree) useful. (n=61)	10% (6)	43% (26)	44% (27)	3% (2)
We found the student admissions and recruitment measures (e.g., program size, GRE scores, work space, health insurance) useful. (n=61)	5% (3)	53% (32)	39% (24)	3% (2)
We found the faculty counts and allocations useful. (n=61)	3% (2)	43% (26)	44% (27)	10% (6)
We found the listings of 18 student activities and campus resources useful. (n=61)	2% (1)	13% (8)	74% (45)	11% (7)

Table A4.9-4 – Economics Program Rankings of NRC Study Use Factors within Program

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My graduate program had active involvement among faculty and staff for the NRC study. (n=58)	2% (1)	28% (16)	53% (31)	17% (10)
My graduate program had one or more champions for the NRC study. (n=58)	4% (2)	10% (6)	64% (37)	22% (13)
My graduate program understood the NRC study methodology and results. (n=57)	12% (7)	37% (21)	46% (26)	5% (3)
My graduate program was persuaded to implement change upon seeing the NRC study results. (n=58)	2% (1)	14% (8)	64% (37)	21% (12)
My graduate program has used or will use the NRC study results to justify decisions. (n=57)	5% (3)	28% (16)	56% (32)	11% (6)
The NRC study broadly has or will improve the quality of my graduate program. (n=58)	3% (2)	26% (15)	54% (31)	17% (10)

Table A4.9-5 – Economics Program Rankings of NRC Study Use Factors within University

Question	Strongly Agree	Agree	Disagree	Strongly Disagree
My campus was persuaded to implement change upon seeing the NRC study results. (n=57)	2% (1)	12% (7)	72% (41)	14% (8)
My campus has used or will use the NRC results to justify decisions. (n=54)	4% (2)	43% (23)	44% (24)	9% (5)
The NRC study broadly has or will improve the quality of graduate education at my campus. (n=57)	2% (1)	19% (11)	61% (35)	18% (10)

REFERENCES

AAU Membership listing website, http://www.aau.edu/about/article.aspx?id=5476.

AAU Association of Graduate Schools Letter to the National Academy of Sciences, January 7, 2011.

AAU Institutional Data Committee Recommendations Implementation Memo, September 28, 2012.

Aldrich, Howard E. and Ruef, Martin. (2006). *Organizations Evolving* (2nd Edition). London, UK: Sage.

Alexander, F. King. (July-August 2000). The Changing Face of Accountability: Monitoring and Assessing Institutional Performance in Higher Education. *The Journal of Higher Education, Vol. 71, No. 4*, 411-431.

Altbach, Philip G. (November 22, 2010). The State of the Rankings. *Inside Higher Ed*.

----. (2012). The Globalization of College and University Rankings. *Change: The Magazine of Higher Learning, Vol. 44, No. 1*, 26-31.

American Mathematics Society, Peter Mucha. (2011). A Quantitative Criticism of the 2010 NRC Assessment of Graduate Programs in Mathematics. Accessed online, http://www.ams.org/profession/data/annual-survey/mucha.pdf.

American Sociological Association. (February 2011). Report to the American Sociological Association Council Regarding the 2010 National Research Council Assessment of Doctorate Programs.

APLU Membership listing website, http://www.aplu.org/page.aspx?pid=249.

Berrett, Dan. (September 18, 2012). All About the Money. *The Chronicle of Higher Education*.

Biglan, Anthony. (1973). The Characteristics of Subject Matter in Different Academic Areas. *Journal of Applied Psychology, Vol. 57, No. 3,* 195-203.

Boyce, Mary E. (Winter 2003). Organizational Learning is Essential to Achieving and Sustaining Change in Higher Education. *Innovative Higher Education, Vol. 28, No. 2,* 119-136.

Breneman, David D. (2005). National Report Card on Higher Education. Public Policy for Academic Quality Research Program. Accessed online, http://www.unc.edu/ppaq/docs/RepCard/RepCard.pdf.

Brewer, D., Gates, S. M., and Goldman, C. A. (2002). *In Pursuit of Prestige: Strategy and Competition in US Higher Education*. New Brunswick, NJ: Transaction Press.

Brooks, Rachelle L. (Fall 2005). Measuring University Quality. *The Review of Higher Education, Vol. 29, No. 1,* 1-21.

Buela-Casal, Gualberto, Gutierrez-Martinez, Olga, Bermudez-Sanchez, Maria, and Vadillo-Munoz, Oscar. (2007). Comparative Study of International Academic Rankings of Universities. *Scientometrics, Vol. 71.3,* 349-365.

Burke, J.C. (Ed.) (2001). *Achieving Accountability in Higher Education*. Baltimore, MD: Johns Hopkins University Press.

Burris, Val. (April 2004). The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks. *American Sociological Review, Vol. 69,* 239-264.

Carnegie Foundation Classification of Institutions of Higher Education, http://classifications.carnegiefoundation.org/.

Cassuto, Leonard. (January 28, 2013). What Are Low-Ranked Graduate Programs Good For? *The Chronicle of Higher Education*.

Cave, M., Hanney, S., Henkel, M., and Kogan, M. (1997). *The Use of Performance Indicators in Higher Education: The Challenge of the Quality Movement* (3rd Edition). London, UK: Jessica Kingsley.

Center for Higher Education Development (CHE) and German Academic Exchange Service University Rankings, https://www.daad.de/deutschland/studienangebote/ranking/en/.

Clark, Burton R. (2003). Sustaining Change in Universities: Continuities in Case Studies and Concepts. *Tertiary Education and Management, Vol. 9,* 99-116.

Cohen, Michael D., March, James G., and Olsen, Johan P. (March 1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, *Vol. 17, No. 1,* 1-25.

Cole, Jonathan R. (April 24, 2011). Too Big to Fail. *The Chronicle of Higher Education*.

Computer Research Association. (September 2010). NRC Doctoral Rankings and Computer Science. Statement Regarding NRC Study Release. Accessed online, http://cra.org/govaffairs/blog/2010/09/nrc-doctoral-rankings-and-computer-science/.

Computing Research Association, by Eric Grimson, Board Chair. (May 2010). Dangers of Rankings with Inaccurate Data. *Computing Research News, Vol. 22, No. 3*.

Computing Research Association, by Eric Grimson, Board Chair. (November 2010). Rating Redux. *Computing Research News, Vol. 22, No. 5*.

Context for Success: Measuring Colleges' Impact. Issue Briefs and Position Papers, accessed online, http://www.hcmstrategists.com/contextforsuccess/.

Council of Graduate Schools, http://www.cgsnet.org/.

Council of Graduate Schools press release. (September 15, 2010). Graduate Education Leaders Issue Global Statement on Quality; Principles and Practices for Assessing the Quality of (Post)-Graduate Education and Research Training. From the Fourth Annual Strategic Leaders Global Summit.

Council of Graduate Schools press release. (September 28, 2010). Statement on Release of NRC Data-Based Assessment of Research-Doctorate Programs.

Council of Gradate Schools. (April 2007). Graduate Education: The Backbone of American Competitiveness and Innovation. Washington, DC.

Dill, David D. (1999). Academic Accountability and University Adaptation: The Architecture of an Academic Learning Organization. *Higher Education, Vol. 38*, 127-154.

----. (September 9, 2006). Convergence and Diversity: The Role and Influence of University Rankings. Keynote address presented at the Consortium of Higher Education Researchers, University of Kassel, Germany.

----. (January 2011). Public Policy Design and University Reform: Insights into Academic Change. Paper presented at the Third International RESUP Conference, Paris, France.

Dill, David D. and Beerkens, Maarja. (Eds.) (2010). *Public Policy for Academic Quality: Analyses of Innovative Policy Instruments*. Dordrecht, The Netherlands: Springer.

Dill, David D. and Soo, Maarja. (2005). Academic Quality, League Tables, and Public Policy: A Cross-National Analysis of University Ranking Systems. *Higher Education, Vol. 49,* 495-533.

DiMaggio, Paul J. and Powell, Walter W. (April 1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review, Vol. 48, No. 2,* 147-160.

Dometrius, Nelson C., Hood III, M.V., Shirkey, Kurt A., and Kidd, Quentin. (December 1998). Bugs in the NRC's Doctoral Program Evaluation Data: From Mites to Hissing Cockroaches. *PS: Political Science and Politics, Vol. 31, No. 4,* 829-835.

Drahl, Carmen. (October 25, 2010). The Un-Rankings: Chemistry graduate programs seek meaning in long-awaited assessment. *Chemical & Engineering News, Vol. 88, No. 43,* 12-17.

Eccles, Charles. (2002). The Use of University Rankings in the United Kingdom. *Higher Education in Europe, Vol. XXVII, No. 4*.

Ehrenberg, Ronald G., Rees, Daniel I., and Brewer, Dominic J. (November 1993). Institutional Responses to Increased External Support for Graduate Students. *The Review of Economics and Statistics, Vol. 75, No. 4,* 671-682.

Ellstrom, Per-Erik. (1983). Four Faces of Educational Organizations. *Higher Education, Vol. 12,* 231-241.

Espeland, Wendy Nelson and Sauder, Michael. (July 2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology, Vol. 113, No. 1,* 1-40.

Ewell, Peter T. (2004). The National Survey of Student Engagement (NSSE). Public Policy for Academic Quality Research Program. Accessed online, http://www.unc.edu/ppaq/docs/NSSE.pdf.

Fain, Paul. (October 22, 2012). Better Measures of College Performance. *Inside Higher Ed*.

Feldman, Maryann, Feller, Irwin, Bercovitz, Janet, and Burton, Richard. (January 2002). Equity and the Technology Transfer Strategies of American Research Universities. *Management Science, Vol. 48, No. 1,* Special Issue on University Entrepreneurship and Technology Transfer, 105-121.

Flaherty, Colleen. (December 17, 2012). Graduate Students and Faculty at U. of Minnesota Pilot New, Ongoing Program Review Process. *Inside Higher Ed*.

Gater, Denise S. (June 2003). Using National Data in University Rankings and Comparisons. *TheCenter* Reports at the University of Florida.

Geiger, Roger and Feller, Irwin. (May-June 1995). The Dispersion of Academic Research in the 1980s. *The Journal of Higher Education, Vol. 66, No. 3,* 336-360.

Gioia, Dennis A. and Thomas, James B. (1996). Identity, Image, and Issue Interpretation: Sensemaking During Strategic Change in Academia. *Administrative Science Quarterly, Vol. 41,* 370-403.

Glenn, David. (June 13, 2010). Doctoral-Program Rankings, Delayed Years, May Be Merely a Historical Record. *The Chronicle of Higher Education*.

----. (September 26, 2010). The NRC Report: Provosts in Electrified Cages. *The Chronicle of Higher Education*.

----. (September 30, 2010). A Critic Sees Deep Problems in the Doctoral Rankings. *The Chronicle of Higher Education*.

----. (December 1, 2010). Measurement of 'Learning Outcomes' Comes to Graduate School. *The Chronicle of Higher Education*.

----. (December 2, 2010). How 3 Graduate Deans are Putting the NRC Rankings to Use. *The Chronicle of Higher Education*.

----. (March 6, 2011). NRC Plans to Release Revised Doctoral-Program Rankings Soon. *The Chronicle of Higher Education*.

Goldstein, Harvey, and Spiegelhalter, David J. (1996). League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society, 159.3,* 385-443.

Gormley, William T. and Weimer, David L. (1999). *Organizational Report Cards*. Cambridge, MA: Harvard University Press.

Grasgreen, Allie. (March 7, 2011). Reconciling with Rankings. *Inside Higher Ed*.

Greger, J.L. (September 2006). The National Research Council's Assessment of Research Doctorate Programs Can Be Used to Strengthen Doctoral Programs in Nutrition. *The Journal of Nutrition, Vol. 136,* 2962-2964.

Grunig, Stephen D. (January-February 1997). Research, Reputation, and Resources: The Effect of Research Activity on Perceptions of Undergraduate Education and Institutional Resource Acquisition. *The Journal of Higher Education*, *Vol. 68, No.1,* 17-52.

Gumport, Patricia J. Graduate Education and Research: Interdependence and Strain. In Altbach, Philip G., Berdahl, Robert O. and Gumport, Patricia J. (Eds.)

(2005). *American higher education in the twenty-first century: social, political, and economic challenges* (pp. 396-426). Baltimore, MD: Johns Hopkins University Press.

Harris, Kerri-Lee, and James, Richmond. (2006). The Course Experience Questionnaire, Graduate Destinations Survey and Learning and Teaching Performance Fund in Australian Higher Education. Public Policy for Academic Quality Research Program. Accessed online, http://www.unc.edu/ppaq/docs/CEQ_final.pdf.

Hattendorf, Lynn C. (Spring 1986). College and University Rankings: An Annotated Bibliography of Analysis, Criticism, and Evaluation. *RQ, Vol. 25, No. 3,* 332-347.

----. (July 15, 1996). Educational Rankings of Higher Education: Fact or Fiction? Paper presented at the Eighth International Conference on Assessing Quality in Higher Education, Queensland, Australia.

Hazelkorn, Ellen. (2007). The Impact of League Tables and Ranking Systems on Higher Education Decision Making. *Higher Education Management and Policy, Vol. 19, No. 2*.

----. (2008). Learning to Live with League Tables and Ranking: The Experience of Institutional Leaders. *Higher Education Policy, Vol. 21*, 193-215.

----. (2009). Rankings and the Battle for World-class Excellence: Institutional Strategies and Policy Choices. *Higher Education Management and Policy, Vol. 21, No. 1*.

----. (2011). *Ranking and the Reshaping of Higher Education: The Battle for World-Class Excellence*. New York, NY: Palgrave Macmillan.

----. (October 3, 2012). Global University Rankings: The New Olympic Sport? *The Chronicle of Higher Education*.

----. (May 23, 2013). Has Higher Education Lost Control Over Quality? *The Chronicle of Higher Education*.

Henry, Gary T. (2000). *Using Evaluation Findings for Policy: A Realist Perspective*. 2000 European Evaluation Society Conference.

----. (2003). Influential Evaluations. *The American Journal of Evaluation, 24.4,* 515-24.

Henry, Gary T., and Mark, Melvin M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation, 24.3,* 293-314.

Hicks, Diana. (March 2008). Evolving Regimes of Multi-university Research Evaluation. Working Paper #27, Georgia Institute of Technology Working Paper Series.

Holmes, Thomas J. (2010). Structural, Experimentalist, and Descriptive Approaches to Empirical Work in Regional Economics. *Journal of Regional Science, Vol. 50, No. 1,* 5-22.

Institute for Higher Education Policy. (May 2009). *Impact of College Rankings on Institutional Decision Making: Four Country Case Studies*. Washington, DC: Institute for Higher Education Policy.

Jaschik, Scott. (May 10, 2010). Methodology Change for Ph.D. Rankings. *Inside Higher Ed*.

----. (September 29, 2010). You're Not No. 1. *Inside Higher Ed*.

----. (March 21, 2011). Sociologists Blast Doctoral Rankings. *Inside Higher Ed*.

----. (April 11, 2011). 'Substantive' Errors in Grad Rankings. *Inside Higher Ed*.

----. (April 15, 2013). QS Changes Rankings Rules Following Recruitment Effort by Irish University. *Inside Higher Ed*.

Kiley, Kevin. (March 13, 2012). ACE Annual Meeting Focuses on Enacting Change Within Shared Governance. *Inside Higher Ed*.

King, Roger, Locke, William, Puncher, Mark, Richardson, John, and Verbik, Line. (2008). *Counting What is Measured or Measuring What Counts? League Tables and Their Impact on Higher Education Institutions in England*. Bristol, UK: Higher Education Funding Council for England.

Kivisto, Jussi. (June 2007). Agency Theory as a Framework for the Government-University Relationship. (Doctoral Dissertation, accessed at: http://tampub.uta.fi/handle/10024/67724).

----. (2008). An Assessment of Agency Theory as a Framework for the Government-University Relationship. *Journal of Higher Education Policy and Management, Vol. 30(4),* 339-350.

Kivisto, Jussi and Holtta, Seppo. (2008). Information as a Regulative Element in Higher Education Systems. *Tertiary Education and Management (TEAM), Vol. 14 (4),* 331-344.

Kuh, Charlotte. (November 2009). The NRC Assessment of Research Doctorate Programs: Thinking Ahead. *Council of Graduate Schools Communicator*, *Vol. 42, No. 9*.

Labi, Aisha. (April 12, 2013). University Rankings Proliferate, Along With New Uses for the Data They Collect. *The Chronicle of Higher Education.*

Lane, Saunders Mac. (August 20, 1996). What Ranks for Research Doctoral Programs? *Proceedings of the National Academy of Sciences of the United States of America, Vol. 93, No. 17,* 8816-8817.

Lederman, Doug. (November 23, 2005). Rating Doctoral Programs. *Inside Higher Ed*.

----. (June 7, 2012). ACE Panel Calls for Sustaining but Changing Regional Accreditation. *Inside Higher Ed*.

Lorden, Joan F, and Martin, Lawrence. (2000). Educational Effectiveness of Doctoral Education. *Journal for Higher Education Strategists, Vol. 1, No. 2,* 191-200.

Lounsbury, Michael. (2001). Institutional Sources of Practice Variation: Staffing College and University Recycling Programs. *Administrative Science Quarterly, Vol. 46,* 29-56.

Maher, Brendan A. (November/December 1996). The NRC's Report on Research Doctorate Programs: Its Uses and Misuses. *Change, Vol. 28, Issue 6,* 54-59.

Mark, Melvin M., Henry, Gary T., and Julnes, George. (1999). Toward an Integrative Framework for Evaluation Practice. *American Journal of Evaluation, 20.2,* 177-98.

Massy, William F. (2001). Academic Audit for Accountability and Improvement. In Burke, J.C. (Ed.) *Achieving Accountability in Higher Education* (pp. 173-197). Baltimore, MD: Johns Hopkins University Press.

----. (January 2013). Initiatives for Containing the Cost of Higher Education. Prepared for the American Enterprise Institute, Washington, DC.

Meister-Scheytt, Claudia and Scheytt, Tobias. (January 2005). The Complexity of Change in Universities. *Higher Education Quarterly, Vol. 59, No. 1,* 76-99.

Merisotis, Jamie P. (2002). Summary Report of the Invitational Roundtable on Statistical Indicators for the Quality Assessment of Higher/Tertiary Education Institutions: Ranking and League Table Methodologies. *Higher Education in Europe*, *Vol. XXVII, No. 4*.

Miller, Arthur H., Tien, Charles, and Peebler, Andrew A. (December 1996). Department Rankings: An Alternative Approach. *PS: Political Science and Politics*, *Vol. 29, No. 4,* 704-717.

Moy, Ernest, Griner, Paul F., Challoner, David R., and Perry, David R. (January 27, 2000). Distribution of Research Awards from the National Institutes of Health among Medical Schools. *New England Journal of Medicine, Vol. 342, No. 4,* 250-255.

National Academies, http://www.nationalacademies.org/.

National Academies and National Research Council. (June 14, 2012). Research Universities and the Future of America. Washington, DC: National Academies Press. Also, press releases and reaction statements from the Association of American Universities as collected by the author.

National Research Council Data-Based Assessment of Research-Doctorate Programs, Ostriker, Jeremiah P., Kuh, Charlotte V., and Voytuk, James A., (Eds.); Committee to Assess Research-Doctorate Programs, Report and Database. Available for download at http://www.nap.edu/rdp/.

National Research Council Data-Based Assessment of Research-Doctorate Programs project website, http://sites.nationalacademies.org/pga/resdoc/index.htm.

National Research Council Data-Based Assessment of Research-Doctorate Programs release webinar. (September 2010). Personal notes and observations.

National Research Council Convocation on Analytic Uses and Future Directions. (March 2011). Personal notes and observations.

O'Leary, John. (February 19, 2013). U-Multirank: In Trouble Already? QS Intelligence Unit, Global Academic Advisory Board. Accessed online, http://www.iu.qs.com/2013/02/19/the-u-multirank-in-trouble-already/.

Olds, Kris and Robertson, Susan. (June 4, 2012). Towards a Global Common Data Set for World University Rankers. *Inside Higher Ed*.

Patton, Michael Quinn. (2000). Utilization-Focused Evaluation. In D.L. Stufflebeam, G.F. Madaus, T. Kellaghan (eds), *Evaluation Models* (pp. 425-438). Boston, MA: Kluwer Academic Publishers.

Pfeffer, Jeffrey and Salancik, Gerry. (1978). *The External Control of Organizations*. Chapters 1-3. Accessed via course posted materials in UNC Blackboard.

Rawlings III, Hunter R. (September 27, 2012). President of the Association of American Universities. Personal observation from remarks provided at UNC-Chapel Hill.

Richardson, Richard C. and Smalling, Thomas R. (2001). Accountability and Governance. In Burke, J.C. (Ed.) *Achieving Accountability in Higher Education* (pp. 55-77). Baltimore, MD: Johns Hopkins University Press.

Russel, William B., Gibeling, Jeffery C., and Weiss, Janet A. (March 2011). Looking Beyond the NRC Assessment of Research Doctorate Programs. *Council of Graduate Schools Communicator*, *Vol. 44, No. 2*.

Rutherford, Desmond, Fleming, William, and Mathias, Haydn. (1985). Strategies for Change in Higher Education: Three Political Models. *Higher Education*, *Vol. 14,* 433-445.

Santiago, Paulo, Tremblay, Karine, Basri, Ester, and Arnal, Elana. (2008). Assuring and Improving Quality. In Tertiary Education for the Knowledge Society: Volume 1, Special Features: Governance, Funding, Quality, Organisation for Economic Co-operation and Development.

Schmidt, Peter. (April 25, 2010). Academics, Too, can be Led Astray by College Rankings, Study Finds. *The Chronicle of Higher Education, 56.33*.

Selingo, Jeff. (March 14, 2013). Let's Kill the Prestige Race Before It Kills Higher Ed. *The Chronicle of Higher Education*.

Shavelson, R.J. (2010). *Measuring College Learning Responsibly: Accountability in a New Era*. Stanford, CA: Stanford University Press.

Sims, Leslie B. and Syverson, Peter D. (2000). Utilizing Data for Effective Administration of Graduate Education. *Journal for Higher Education Strategists, Vol. 1, No. 1,* 65-94.

Sine, Wesley David, Shane, Scott, and Di Gregorio, Dante. (April 2003). The Halo Effect and Technology Licensing: The Influence of Institutional Prestige on the Licensing of University Inventions. *Management Science, Vol. 49, No. 4,* Special Issue on Managing Knowledge in Organizations: Creating, Retaining, and Transferring Knowledge, 478-496.

Soo, Maarja and Dill, David D. (2007). The CHE University Ranking in Germany. Public Policy for Academic Quality Research Program. Accessed online, http://www.unc.edu/ppaq/docs/CHE_formatted.pdf.

Srikanthan, Gitachari and Dalrymple, John F. (2007). A Conceptual Overview of a Holistic Model for Quality in Higher Education. *International Journal of Educational Management, Vol. 21, No. 3,* 173-193.

Stigler, Stephen M. (August 20, 1996). Rating, not Ranking. *Proceedings of the National Academy of Sciences of the United States of America, Vol. 93, No. 17,* 8818-8819.

Suskie, Linda. (October 26, 2010). Why Are We Assessing? *Inside Higher Ed*.

Tan, David L. (1986). The Assessment of Quality in Higher Education: A Critical Review of the Literature and Research. *Research in Higher Education, Vol. 24, No. 3*, 223-265.

The Times Higher Education World University Rankings, http://www.timeshighereducation.co.uk/world-university-rankings/.

Thomson Reuters. (March 19, 2013). Thomson Reuters Launches Metrics Mania to Crown Research Champion. Press Release/Product Overview.

Toutkoushian, Robert K, Dundar, Halil, and Becker, William E. (1998). The National Research Council Graduate Program Ratings: What Are They Measuring? *The Review of Higher Education, 21.4,* 427-443.

Trounson, Andrew. (February 13, 2013). Australia Adapts European Union Tool for Comparing Universities. *Inside Higher Ed*.

----. (March 20, 2013). Australian Universities Dedicate Positions to Working with Rankings Groups. *Inside Higher Ed*.

Trow, Martin. (1999). Biology at Berkeley: A Case Study of Reorganization and its Costs and Benefits. *Research and Occasional Paper Series, CSHE.1.99*.

University websites and press releases, accessed September 28-30, 2010 (Boston University; University of Michigan; Cornell University; Duke University; University of California-Los Angeles; University of California-Berkeley; Boston University; University of Virginia).

Usher, Alex. (March 2013). Blog accessed at http://higheredstrategy.com/the-paradox-of-university-rankings/, and as referenced in: Olds, Kris. (March 4, 2013). On the Illogics of the Times Higher Education Reputation Rankings. *Inside Higher Ed*.

Van Dyke, Nina. (July 2005). Twenty Years of University Report Cards. *Higher Education in Europe, Vol. 30, No. 2*.

Van Dooren, Wouter. (August 2005). What Makes Organisations Measure? Hypotheses on the Causes and Conditions for Performance Measurement. *Financial Accountability & Management, Vol. 21, No. 3,* 363-383.

van Vught, Frans A. and Ziegele, Frank (Eds.) (2012). *Multidimensional Ranking: The Design and Development of U-Multirank*. Dordrecht, The Netherlands: Springer.

Vaughn, John. (2002). Accreditation, Commercial Rankings, and New Approaches to Assessing the Quality of University Research and Education Programmes in the United States. *Higher Education in Europe*, *Vol. XXVII, No. 4*.

Volkwein, J. Fredericks and Grunig, Stephen D. (2001). Resources and Reputation in Higher Education: Double, Double, Toil and Trouble. In Burke, J.C. (Ed.) *Achieving Accountability in Higher Education* (pp. 246-274). Baltimore, MD: Johns Hopkins University Press.

Weiss, Carol H. (1977). *Using Social Research in Public Policy Making*. Lexington, MA: D.C. Heath and Company.

----. (September/October 1979). The Many Meanings of Research Utilization. *Public Administration Review, Vol. 39, No. 5,* 426-431.

Weiss, Robert. (1994). *Learning From Strangers: The Art and Method of Qualitative Interview Studies*. New York, NY: The Free Press.

Wiley, John D. (March 2009). Quality, Accreditation, and Graduate Education: What Does the Future Hold? *Council of Graduate Schools Communicator*, *Vol. 42, No. 2*. Approximate recreation of remarks at the CGS Annual Meeting, December, 2008.

Winston, Gordon C. (December 2001). Higher Education's Costs, Prices, and Subsidies: Some Economic Facts and Fundamentals. National Center for Education Statistics, Statistical Analysis Report, Volume 2: Commissioned Papers, 117-128.

Wellman, Jane. (2003). Do Federal Spending and Regulation Produce 'Quality' Higher Education? The Federal Role in Accountability for Quality in Higher Education. In Thomas R. Wolanin, *Reauthorizing the Higher Education Act (HEA): Issues and Options* (pp. 129-146). Washington, DC: Institute for Higher Education Policy.

Yorke, Mantz and Longden, Bernard. (2005). Significant Figures: Performance Indicators and 'League Tables'. Standing Conference of Principals (SCOP), London.